



Etude fonctionnelle et structurale du régulateur floral LEAFY d'*Arabidopsis thaliana*

Cécile Hamès

► To cite this version:

Cécile Hamès. Etude fonctionnelle et structurale du régulateur floral LEAFY d'*Arabidopsis thaliana*.
Biologie végétale. Université Joseph-Fourier - Grenoble I, 2008. Français. NNT: . tel-00492019

HAL Id: tel-00492019

<https://theses.hal.science/tel-00492019>

Submitted on 22 Jul 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**UNIVERSITE JOSEPH FOURIER
GRENOBLE
ECOLE DOCTORALE CHIMIE ET SCIENCES DU VIVANT**

THÈSE DE DOCTORAT

Présentée par
Cécile HAMES

**Etude fonctionnelle et structurale du régulateur
floral LEAFY d'*Arabidopsis thaliana***

Soutenue le 26 septembre 2008 devant le jury composé de :

Pr. Eva PEBAY-PEYROULA	Présidente
Dr. Mohammed BENDAHMANE	Rapporteur
Dr. Patrick LAUFS	Rapporteur
Dr. Jean-François BRIAT	Examineur
Dr. Carlo PETOSA	Examineur
Dr. François PARCY	Directeur de thèse

Sommaire

Listes des abréviations	7-8
INTRODUCTION	9-54
I. La floraison chez les Angiospermes	9
1. Cycle de vie d'une plante à fleur ou Angiosperme	9
2. Qu'est ce qui déclenche la transition florale ?	11
a) Comment la plante traduit-elle les signaux environnementaux ?	12
b) Les intégrateurs floraux: <i>LFY</i> , <i>FT</i> et <i>SOC1</i>	17
3. Morphogenèse florale	20
a) Les gènes d'identité des méristèmes floraux ou gènes FMI	20
b) Bases du développement de la fleur	23
c) Du modèle ABC au modèle du quartet floral (ou ABCE)	23
II. <i>LFY</i>, un régulateur clé du développement floral	26
1. Présentation	26
a) Description des plantes mutantes <i>lfy</i>	27
b) Description des plantes surexprimant constitutivement <i>LFY</i> (<i>35S::LFY</i>)	28
c) Profil d'expression de <i>LFY</i>	28
2. Rôle dans l'identité méristématique florale	29
a) Comment agit <i>LFY</i> en qualité de FMI?	29
b) Une action partiellement redondante à celle d' <i>AP1</i>	29
c) Activation de gènes FMI	30
d) Répression de gènes d'identité du méristème d'inflorescence	31
3. Rôle de FMI tardif dans la morphogenèse florale	32
a) Mise en évidence	32
b) Rôle de <i>LFY</i> dans l'activation d' <i>AP1</i> , gène 'A' d'identité des sépales et pétales	33
c) Rôle dans l'activation d' <i>AP3</i> et <i>PI</i> , gènes 'B' d'identité des pétales et étamines	33
d) Rôle dans l'activation d' <i>AG</i> , gènes 'C' d'identité des étamines et carpelles	34
e) Modèle de délimitation du profil d'expression des gènes d'identité des organes floraux	35
4. Un rôle clé généralisable aux Angiospermes	35
III. <i>LFY</i>, un facteur de transcription original	36
1. Généralités sur les facteurs de transcription	36
a) Les grandes familles de protéines associées à l'ADN	37
b) Etat des connaissances sur les facteurs de transcription végétaux	40
2. Points sur les connaissances biochimiques relatives à <i>LFY</i>	41
a) Régulation de l'expression d' <i>AP1</i>	42
b) Régulation de l'expression d' <i>AG</i>	42
c) Régulation de l'expression d' <i>AP3</i>	45
d) Evolution de <i>LFY</i>	46
IV. <i>LFY</i>, un acteur central de l'apparition des fleurs ?	47
1. « Un abominable mystère »	47
2. Facteurs de transcription d'intérêt	49
3. Théories sur l'apparition des fleurs	50
a) <i>Mostly Male</i>	50
b) <i>Out of Male/Out of Female</i>	51
c) Premières caractérisations évolutives de la protéine <i>LFY</i>	51
V. Objectifs de mon travail de thèse	53
MATERIEL	55
1. Matériel bactérien	55
2. Levure	55
3. Matériel végétal	55
METHODES	55-68
1. Solutions d'usage courant	55
2. Biologie moléculaire	56
a) BiFC	56
b) Vecteurs d'expression en bactéries	56

c) Vecteurs d'expression chez la levure	58
3. Expression et purification des protéines recombinantes	59
Protocole d'induction alternatif des protéines LFY-C	59
Protocole d'induction de LFY et WUS	59
Protocole de purification de LFY FL	59
Protocole de purification de WUS FL et WUS Hd	60
Protocole de purification de la protéase TEV	60
4. Etude biochimique des protéines	60
a) Gel d'acrylamide	60
b) Western Blot contre la protéine WUS	61
c) Dosage de protéines	61
d) Anisotropie de fluorescence (AF) et expériences de retard sur gel (EMSA)	62
e) Technique de SEC (<i>Size Exclusion Chromatography</i>) analytique	63
f) Technique de MALLS-SEC (<i>Multi-Angle Laser Light Scattering</i>)	64
g) Technique de SPR (<i>Surface Plasmon Resonance</i>)	64
5. Cristallo-genèse	64
a) LFY-C/ADN	64
b) LFY-C	65
6. Caractérisation de LFY chez la levure	65
a) Milieux de culture	65
b) Transformation des levures	65
c) Protocole de mesure de l'activité β -galactosidase	66
7. Caractérisation de LFY chez les plantes	67
a) Stérilisation des graines	67
b) Conditions de culture des plantes	67
c) Expression transitoire en feuilles de tabac <i>N.benthamiana</i>	67
d) Transformation stable d' <i>A. thaliana</i> par <i>A. tumefaciens</i>	67
e) Microscopie confocale	68
f) Extraction de l'ADN génomique (ADNg) des plantes d' <i>A. thaliana</i>	68
g) Génotypage des plantes <i>lfy-12</i>	68
RESULTATS CHAPITRE 1	69-94
I. Caractérisation biochimique du domaine conservé LFY-C	69
1. Introduction: pourquoi étudier le domaine C-terminal ?	69
2. Production du domaine C-terminal de LFY: sélection du vecteur d'expression optimal	69
3. Fonctionnalité et caractérisation quantitative de la liaison de LFY-C à l'ADN	72
a) Principe des techniques	73
c) Estimation de la constante d'affinité de LFY pour des régions régulatrices entières par SPR (travail d'E.Moyroud)	76
d) Influence de la longueur de l'oligonucléotide sur l'affinité de liaison de LFY-C à l'ADN	76
4. Domaine minimal de liaison de LFY à l'ADN	77
5. Détermination de la nature des deux complexes de LFY-C/ADN	78
a) Ferguson Plot	79
b) EMSA sur gel gradient à l'équilibre (GEMSA)	80
c) Chromatographie par exclusion de taille (SEC) analytique	80
d) Premier bilan: LFY se lierait à l'ADN selon un ratio 4LFY-C/ADN ?	81
e) Ultracentrifugation analytique (AUC; en collaboration avec Christine EBEL, IBS, Grenoble)	81
f) MALLS-SEC (en collaboration avec Francine Gérard et Marc Jamin, IVMS, Grenoble)	82
g) EMSA en présence d'un mélange équimolaire de LFY-C de masses différentes	82
h) Deuxième bilan: LFY se lie à l'ADN selon un ratio 2 LFY-C/ADN	83
6. Mise en évidence d'une coopérativité de liaison de LFY-C à l'ADN	84
a) Définition de la coopérativité	84
b) Evaluation de la coopérativité de LFY-C	85
c) Influence de la séquence d'ADN sur la coopérativité	85
7. Etude de la coordination d'ion(s) métallique(s)	85
a) Les indices à l'origine de cette hypothèse	85
b) Résultats préliminaires par EMSA et AF	86
c) Analyses fines par dosage des métaux	89
8. Discussion	90
a) Ce que l'approche biochimique nous a apporté	90
b) Ce à quoi l'approche biochimique ne permet pas de répondre	92
c) Réflexions sur nos résultats	92

RESULTATS CHAPITRE 2 **95-110**

II. Etude structurale du domaine conservé C-terminal de LFY	95
1. Protocole de cristallogénèse	95
2. Cristallographie	97
a) Structure globale	97
b) Caractéristiques du domaine de liaison à l'ADN de LFY-C	98
c) Similarités structurales	99
d) Vers la compréhension du mécanisme moléculaire de la coopérativité	101
3. Validation et approfondissement des données structurales par approche biochimique	101
a) Validation des points de contact LFY-C/AP1	101
b) Etude du comportement de liaison des mutants connus de LFY <i>in planta</i>	103
c) Confirmation des bases moléculaires à l'origine de la coopérativité	104
4. Discussion	105
a) Un nouveau domaine de liaison à l'ADN	105
b) Un domaine extrêmement conservé de la Mousse aux végétaux supérieurs	105
c) Identification des points de contact LFY-C/ADN	105
d) Bases moléculaires de la coopérativité	106
e) Par quelle voie LFY-C lie-t-elle l'ADN?	107
f) Un complexe d'ordre supérieur ?	109

RESULTATS CHAPITRE 3 **111-130**

III. Caractérisation fonctionnelle de la protéine LFY FL	111
1. Introduction	111
2. Etude <i>in vitro</i> du comportement de liaison à l'ADN de LFY FL	111
a) Une protéine difficile à produire	111
b) Premières caractérisations fonctionnelles	112
3. Etude de l'homodimérisation de LFY FL <i>in vivo</i> chez la levure	117
a) Mise en évidence d'une dimérisation de LFY sur l'ADN par technique de double hybride	117
b) Mise en évidence d'une implication des résidus His387 et Arg390 dans la capacité de LFY FL à activer AG par simple hybride (travail d'E. Thévenon)	119
4. Etude de l'homodimérisation de LFY FL <i>in planta</i>	119
a) Mise en évidence d'une dimérisation de LFY chez <i>A. thaliana</i>	120
b) Mise en évidence d'une oligomérisation partiellement nucléaire par BiFC	122
c) Mise en évidence d'une implication des résidus His387 et Arg390 dans la capacité de LFY FL à activer AP1 par expression transitoire chez le tabac (travail d'E. Thévenon)	125
5. Discussion	126
a) Ce que l'approche biochimique nous a apporté	126
b) Ce que les expériences <i>in vivo</i> nous ont apporté	129

DISCUSSION GENERALE **131-154**

Bilan	131
1. Les données structurales sur le repliement de la protéine LFY	131
a) Un facteur de transcription original de la superfamille des protéines HTH	131
b) Quelle est l'origine de LFY ?	132
c) Une haute conservation de la protéine	132
d) L'ADN intervient-il sur la conformation de LFY ?	132
2. La coopérativité de liaison de LFY-C à l'ADN a-t-elle une réalité biologique?	133
a) La coopérativité de liaison de LFY-C est-elle valable dans le contexte LFY FL ?	133
b) La coopérativité intervient-elle dans l'activité de LFY FL <i>in vivo</i> ?	134
c) Quelle est la signification des complexes LFY/ADN d'ordre supérieur ?	135
3. Caractérisation moléculaire de l'interaction LFY-C/ADN	137
a) Détermination de l'affinité spécifique de liaison de LFY à l'ADN	137
b) Quelles séquences d'ADN sont reconnues par LFY ?	137
4. Comment LFY agit-il avec ses corégulateurs ?	139
a) Cas de la régulation d'AG par LFY et WUS	139
b) Cas de la régulation d'AP3 par LFY et UFO	145
c) A la recherche de nouveaux corégulateurs	147
5. Comment LFY a-t-il évolué ?	147
6. Quelle relation existe-il entre LFY et l'apparition des plantes à fleurs ?	152

Références bibliographiques	155-162
Annexe 1	163-184
Annexe 2	185-193

Listes des abréviations

35S: promoteur constitutif fort de l'ARN 35S du virus de la mosaïque du chou-fleur
aa: acides aminés
AD: Activation Domain, domaine d'activation de la transcription
ADN: Acide DésoxyriboNucléique
AF: Anisotropie de Fluorescence
AG: AGAMOUS
AGL19/24: AGAMOUS-LIKE19, AGAMOUS-LIKE24
ANA: Amborellaceae-Nymphaeales-Austrobaileyales, Angiospermes basales
AP1/3: APETALA1, APETALA2, APETALA3
ARN: Acide RiboNucléique
AS: Acide Salicylique
AUC: Analytical UltraCentrifugation, ultracentrifugation analytique
BiFC: Bimolecular Fluorescence Complementation
C1: complexe minoritaire formé par une protéine LFY-C sur l'ADN
C2: complexe majoritaire formé par deux protéine LFY-C sur l'ADN
CAL: CAULIFLOWER
ChIP: Chromatine ImmunoPréciptation
CO: CONSTANS
Col: Columbia
DBD: DNA-Binding domain, domaine de liaison à l'ADN
DEF: DEFICIENS, homologue d'AP3 chez *A. majus*
DLS: Dynamic Light Scattering
EDTA: Acide Ethylène-Diamine-Tétraacétique
EFS: EARLY FLOWERING IN SHORT DAYS
EMS: Ethyl MéthaneSulfonate
EMSA: Electrophoretic Mobility Shift Assay, expérience de retard sur gel
ESD4: EARLY IN SHORT DAYS4
FD: FLOWERING LOCUS D
FIM: FIMBRIATA, homologue d'UFO chez *A. majus*
FLC: FLOWERING LOCUS C
FLO: FLORICAULA
FMI: Floral Meristem Identify
FPF1: FLOWERING PROMOTIVE FACTOR1
FRET: Fluorescent Resonance Energy Transfer, transfert d'énergie de fluorescence par résonance
FRI: FRIGIDA
FT: FLOWERING LOCUS T
FUL: FRUITFULL
GAs: Gibbérellines
GAI: GIBBERELLIC ACID INSENSITIVE
GEMSA: Gradient Electrophoretic Mobility Shift Assay, retard sur gel gradient à l'équilibre
GFP: Green Fluorescent Protein, protéine fluorescente verte
GLO: GLOBOSA, homologue de PI chez *A. majus*
GR: récepteur aux glucocorticoïdes du rat
GST: Gluthatione S-Transferase
HPR: Peroxydase de raifort
HTH: Helix-Turn-Helix, motif de liaison à l'ADN de type hélice-tour-hélice
ICP-MS: Inductively Coupled Plasma-Mass Spectrometer, spectrométrie de masse couplée à un plasma inductif
IPTG: IsoPropyl-beta-Thio-Galactoside
k: constante de vitesse de dissociation
K_d: constante de dissociation
kDa: kiloDalton, unité de mesure de masse
LD: LUMINIDEPENDENS
LFY FL: LEAFY Full Length, protéine LFY pleine longueur
LFY: LEAFY
LFY-C: domaine C-terminal de LFY délimité par les résidus 223-424
LFY-DBD: domaine de la région conservé C-terminal de LFY impliquée dans la liaison à l'ADN
MADS: protéine MCM1, AGAMOUS, DEFICIENS, SRF
MALLS: Multi-Angle Laser Light Scattering, mesure de diffusion statique de la lumière selon un angle de 360°
microPIXE: micro Proton Induced X-ray Emission
MW, Molecular Weight, poids moléculaire

NLY: NEEDLY, paralogue de LFY chez les Gymnospermes
 NO: Nitric Oxyde, Oxyde nitrique
 PEBP: PhosphatidylEthanolamine Binding Protein, Protéine de liaison du phosphatidylethanolamine
 PEG: PolyEthylène Glycol
 PHOR1: PHOTOPERIOD RESPONSIVE1
 PI: PISTILLATA
 PIE1: PHOTOPERIOD INDEPENDENT EARLY FLOWERING1
 PSA: PerSulfate d'Ammonium
 RGA: REPRESSOR OF GA1-3
 RGL1: RGA-LIKE1
 RKIP: Raf Kinase Inhibitor Protein, Protéine inhibitrice de Raf kinase
 SCF: protéine Skp1–Cdc53–F-box
 SDS-PAGE: Sodium DodecylSulfate PolyAcrylamide Gel Electrophoresis, gel de polyacrymide en conditions dénaturantes
 SEC: Size Exclusion Chromatography, chromatographie d'exclusion de taille
 selex: Systematic Evolution of Ligands by Exponential Enrichment, sélection de sites de liaison par enrichissement exponentiel
 SEP1/4: SEPALLATA1, SEPALLATA2, SEPALLATA3, SEPALLATA4
 SHI: SHORT INTERNODES
 SOC1/AGL20: SUPPRESSOR OF OVEREXPRESSION OF CO1/AGAMOUS-LIKE20
 SPR: Surface Plasmon Resonance, résonance plasmonique de surface
 SPY: SPINDLY
 SVP: SHORT VEGETATIVE PHASE
 TEV: Protéase du Virus Tobacco Etch
 TFL1: TERMINAL FLOWER1
 TNT: système de transcription traduction *in vitro*
 TPM: Tethered Particle Motion
 u.a.: Unité Arbitraire
 UAS: Upstream Activator Sequences
 UFO: UNUSUAL FLORAL ORGANS
 VIN3: VERNALIZATION INSENSITIVE3
 VIP: VERNALIZATION INDEPENDENCE
 VP16: facteur de transcription du virus de l'Herpès possédant un domaine activateur constitutif de la transcription
 VRN1/2: VERNALIZATION1, VERNALIZATION2
 WUS: WUSCHEL
 YFP: Yellow Fluorescent Protein, protéine fluorescente jaune
 YFP^C: moitié C-terminale de la YFP utilisée dans l'expérience de BiFC
 YFP^N: moitié N-terminale de la YFP utilisée dans l'expérience de BiFC

I. La floraison chez les Angiospermes

1. Cycle de vie d'une plante à fleur ou Angiosperme

La fleur fait souvent l'objet de convoitise, incarnant la beauté de la nature dans toute sa splendeur et sa diversité. Mais au-delà de ses aspects esthétiques, la fleur a joué un rôle sélectif fulgurant au sein du règne végétal. Apparues sur Terre il y a 120 millions d'années, soit près de 300 millions d'années après l'apparition des premières plantes terrestres (Mousses), les plantes à fleurs ou Angiospermes dominent aujourd'hui très largement ce règne avec plus de 250000 espèces. Le caractère inédit de la fleur, concentration en une même structure des organes reproducteurs mâle (étamines) et femelle (pistil) entourés par deux couronnes externes, les pétales et les sépales, est à l'origine de ce succès. Cette structure facilite la reproduction sexuée par pollinisation, d'autant plus que les pétales colorés et souvent parfumés et la sécrétion de nectar jouent d'un fort pouvoir attracteur sur les insectes pollinisateurs. Cette interaction, mettant souvent en jeu un couple insecte/famille végétale spécifique, a favorisé l'isolement reproductif et par conséquent la spéciation (Feild, 2002; Grant, 1994).

Suite à la pollinisation, l'ovule fécondé donne naissance à une graine qui assure protection et nutrition de l'embryon, et qui est confinée dans un fruit tel une silique dans le cas d'*Arabidopsis thaliana*. La graine peut ensuite entrer en dormance sur des périodes relativement longues (en fonction de l'espèce) afin d'attendre le moment le plus opportun pour germer et garantir la pérennisation de l'espèce. Dès le développement précoce de l'embryon, deux réservoirs de cellules souches se mettent en place: le méristème racinaire est à l'origine de tous les types cellulaires nécessaires à la genèse des racines, et la gemmule ou méristème apical assure le développement aérien de la plante (Fig.1).

Mon travail de thèse a porté sur plante modèle *Arabidopsis thaliana* chez laquelle de nombreuses études génétiques de la floraison et du développement floral ont été accomplies. Cette plante de la famille des *Brassicacées* est très étudiée par la communauté scientifique pour diverses raisons. Elle possède un génome de petite taille qui a été entièrement séquencé en l'an 2000 (The Arabidopsis Genome Initiative, 2000), révélant l'existence de près de 26000 gènes. Considérée comme une 'mauvaise herbe', elle est facile à cultiver et à manipuler génétiquement, elle se développe vite (cycle de développement de 2 mois), est peu encombrante, s'autoféconde, et est très prolifique.

Chez *A. thaliana*, le méristème apical produit de façon indéterminée des tiges et des feuilles qui adoptent une organisation en rosette selon une phyllotaxie spiralée. Les entrenoeuds sont courts. Puis, suite à l'induction florale, le méristème « végétatif » devient « d'inflorescence »: les entrenoeuds s'allongent pour laisser émerger une longue tige, c'est la montaison. A la base de cette inflorescence primaire se trouvent quelques inflorescences secondaires ou coflorescences poussant chacune à l'aisselle d'une feuille. Au dessus de ces coflorescences, la tige porte des fleurs arrangées selon une phyllotaxie spiralée qui contrairement aux coflorescences ne sont pas associées à une feuille modifiée

ou bractée. Ces fleurs constituées de différents organes en nombre déterminé naissent sur les flancs du méristème apical à partir de jeunes bourgeons appelés méristèmes floraux (Fig.1).

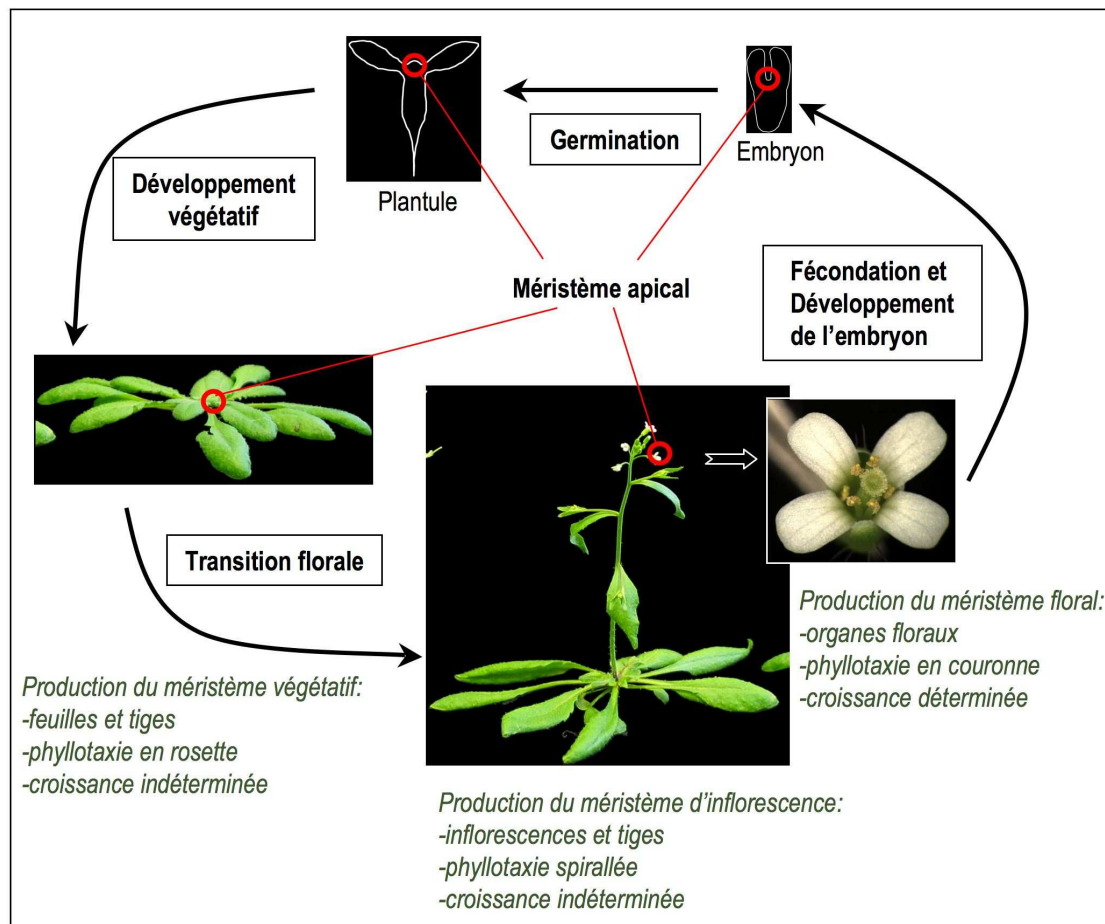


Figure 1: Cycle de vie d'*Arabidopsis thaliana* (adapté d'après Alexis Maizel)

Au cours de son développement, la plante se retrouve confrontée à deux étapes critiques: quand doit-elle germer; et quand doit-elle fleurir ? Dans les deux situations les mécanismes génétiques mis en jeu doivent être suffisamment élaborés pour assurer la meilleure adaptation de la plante à son environnement et ainsi garantir sa survie. Dans le cadre de ma thèse, je n'aborderai que le cas de la transition florale, transition au cours de laquelle le méristème apical passe d'un stade végétatif à un stade d'inflorescence (Fig.1).

Le développement des fleurs est le résultat d'un processus qui peut être divisé en trois phases successives: la transition florale, la mise en place des méristèmes floraux et enfin le développement des organes floraux (Fig.2). Ces trois phases sont sous le contrôle respectif de gènes dit de « temps de floraison », « d'identité méristématique florale » et « d'identité des organes floraux » (Tab.1). Mon sujet d'étude, le régulateur *LFY*, fait à la fois partie des gènes de contrôle du temps de floraison et des gènes d'identité méristématique florale. Dans cette introduction, je décrirai dans un premier temps ces trois phases ainsi que les principaux

régulateurs impliqués sans m'attarder toutefois sur *LFY* que je traiterai de façon plus exhaustive ultérieurement.

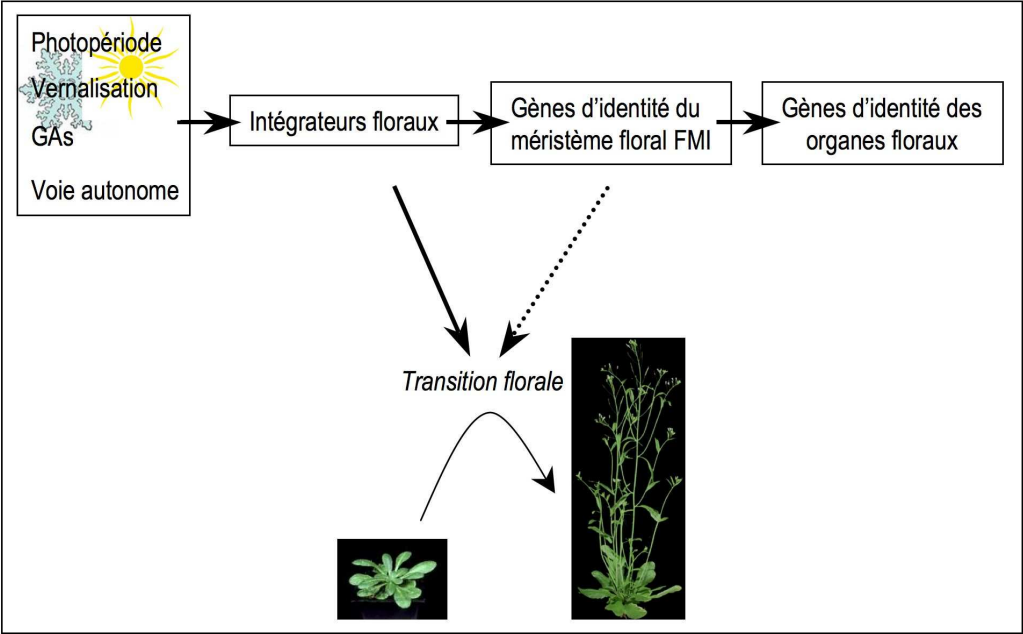


Figure 2: Les différentes étapes de la floraison
 La transition florale est gouvernée essentiellement par les intégrateurs floraux. Les gènes FMI peuvent intervenir sur le temps de floraison (en particulier *LFY*), mais ils sont surtout responsable du changement d'identité du méristème d'inflorescence en méristème floral. Les gènes d'identité des organes floraux interviennent en dernier lieu pour former les différents organes caractéristiques de la fleur.

Tableau 1: Récapitulatif des gènes identifiés à ce jour en tant qu'acteurs du processus de la floraison

Fonction	Gènes de temps de floraison (liste non exhaustive)	Intégrateur floral	Gènes d'identité du méristème floral	Gènes d'identité des organes floraux
Gène	<i>FLC, FRI, VIN3, VRN1, VRN2, FCA, FY, FVE, FPA, LD, FLD, FLK, PIE1, EFS, ESD4, VIP, CO, CRY2, PHYA, GAI, RGA, FPF1, RGL1, PHOR1, SHI, GA1, SPY, MAF1 à MAF5, SVP, TFL1, FT, LFY, SOC1</i>	<i>LFY, FT, SOC1</i>	<i>LFY, AP1, CAL, FUL, UFO, TFL1</i>	<i>AP1, AP2, AP3, PI, AG</i>

2. Qu'est ce qui déclenche la transition florale ?

Le choix du moment le plus propice à la floraison est critique car il conditionne la formation des graines. Toutes les conditions doivent donc être réunies pour assurer le succès préalable de la fécondation. Pour ce faire, beaucoup de plantes gardent en mémoire une période prolongée d'exposition au froid (**vernalisation**) signalant le passage de l'hiver, et perçoivent les changements de

la durée du jour (**photopériode**), indicateurs fiables de la progression saisonnière (Fig.2). Certaines plantes sont alors dites 'de jours courts' car elles sont sensibles à une exposition quotidienne à la lumière n'excédant pas une durée déterminée, d'autres sont dites 'de jours longs' car elles sont au contraire stimulées à partir d'une durée minimale d'exposition lumineuse. D'autres signaux environnementaux influent également, souvent dans une moindre mesure, comme la disponibilité des ressources nutritives (eau, sels minéraux, sucres, azote, CO₂), la température ambiante, la qualité de la lumière, les stress biotiques (plantes avoisinantes, virus, organismes phytophages et pathogènes) et abiotiques (UV, stress salin, vent).

A ces facteurs exogènes s'ajoutent des facteurs endogènes. Plusieurs régulateurs du développement de la plante peuvent en effet intervenir sur le temps de floraison, comme l'acide salicylique, ou des phytohormones telles que les gibbérellines (GAs), l'acide abscissique, l'éthylène, et les brassinostéroïdes (Fig.2). Les GAs ont l'effet le plus drastique: sous exposition à une photopériode défavorable, c'est-à-dire en jours courts dans le cas d'*A. thaliana* (plante de jours longs facultative), les plantes *A. thaliana* mutées dans la voie de biosynthèse des GAs ou insensibles aux GAs sont fortement retardées dans leur floraison, voire incapables de fleurir (Wilson et al., 1992).

a) Comment la plante traduit-elle les signaux environnementaux ?

La transition florale est soudaine et irréversible, reflétant l'activité d'un réseau génétique robuste et efficace. De nombreux gènes impliqués dans ce réseau ont été identifiés depuis la fin des années 80 soit par analyse des variations naturelles du temps de floraison entre les différents écotypes d'*Arabidopsis thaliana* (ex.: *Columbia* (Col), *Landsberg erecta* (Ler) et *Wassilewskija* (Ws); Simpson et al., 1999; Sung and Amasino, 2004), soit par mutagenèse et crible phénotypique. Ces travaux révèlent l'existence de nombreux gènes jouant un effet répresseur sur le temps de floraison, et de quatre grandes cascades d'activation correspondant aux voies dites de la vernalisation, de la photopériode, des GAs et à la voie autonome, dont la finalité est dans tous les cas la régulation des **intégrateurs floraux** *LEAFY* (*LFY*; Blazquez et al., 1997; Nilsson et al., 1998), *FLOWERING LOCUS T* (*FT*; Kardailsky et al., 1999; Kobayashi et al., 1999; Koornneef et al., 1991) et *SUPPRESSOR OF OVEREXPRESSION OF CO1/AGAMOUS-LIKE20* (*SOC1/AGL20*; Onouchi et al., 2000; Samach et al., 2000).

Je commencerai ici par décrire les principaux acteurs de la répression de la floraison pour me consacrer ensuite aux quatre cascades d'activation. Ces descriptions sont transcrites dans un schéma récapitulatif reporté en figure 3 (se reporter également au tableau 1 pour un récapitulatif des gènes de temps de floraison).

1-Le système répressif

Par défaut, la floraison est activement réprimée. Au cours du développement végétatif, la plante doit accumuler des réserves pour survivre lors des processus de floraison et de production de la graine, processus très coûteux en énergie. Il existe donc un ensemble de gènes dont le rôle est d'inhiber la floraison.

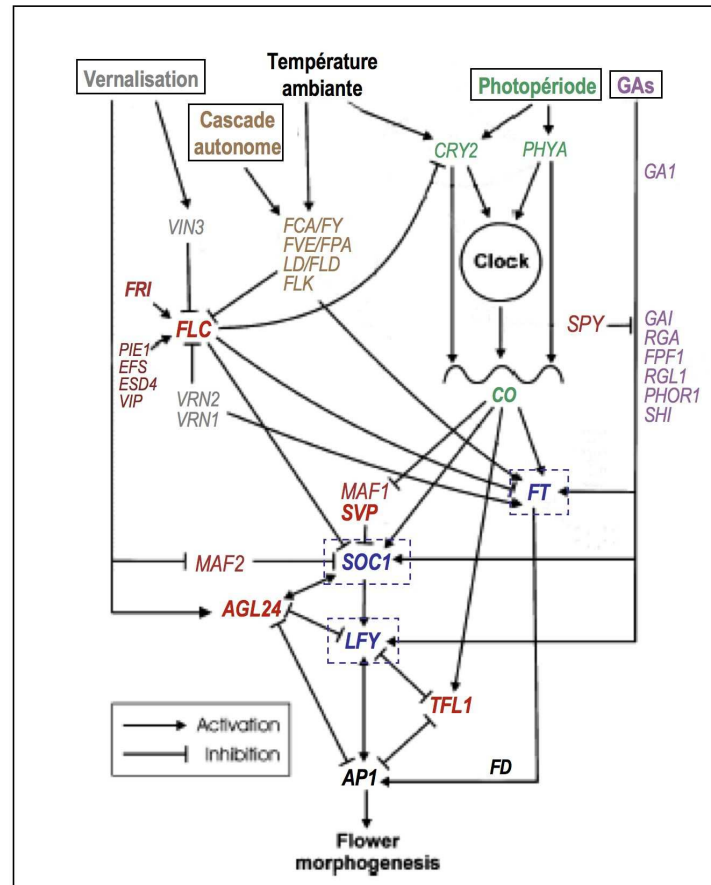


Figure 3: Interaction entre les gènes activant ou réprimant la floraison chez *Arabidopsis thaliana* (adapté à partir de Bernier et Périlleux 2005)

Les principaux signaux de la floraison sont encadrés en noir. Les intégrateurs floraux sont encadrés en bleu. Les répresseurs de la floraison sont signalés en rouge, les acteurs de la voie autonome en marron, de la voie de la photopériode en vert, de la voie de la vernalisation en gris, de la voie des GAs en violet.

Le plus étudié et sans doute le plus important est *FLOWERING LOCUS C (FLC)*, gène codant pour un facteur de transcription à boîte MADS (Michaels and Amasino, 1999). C'est le principal répresseur de la floraison, qui de par sa position en aval des principales voies de signalisation (décrites dans les paragraphes suivants) est extrêmement influent. *FLC* réprime *SOC1*, *FT*, probablement *LFY*, ainsi que le gène *FD* impliqué dans la cascade photopériode (Fig.3; Bernier and Périlleux, 2005; Borner et al., 2000; Helliwell et al., 2006; Hepworth et al., 2002; Lee et al., 2000; Michaels and Amasino, 1999; Nilsson et al., 1998; Samach et al., 2000; Searle et al., 2006). D'autres inhibiteurs de la floraison agissent en activant *FLC* (Fig.3). C'est le cas en particulier de *FRIGIDA (FRI)* (Boss et al., 2004; Clarke and Dean, 1994). Les variations alléliques de *FLC* et *FRI*, qui modulent le besoin en vernalisation,

expliquent à elles seules l'adaptation des différents écotypes d'*Arabidopsis thaliana* à une grande plage de latitudes (Simpson et al., 1999).

En plus de *FLC*, il existe d'autres facteurs qui inhibent la floraison comme *TERMINAL FLOWER1* (*TFL1*; Ruiz-Garcia et al., 1997), et *SHORT VEGETATIVE PHASE* (*SVP*; Hartmann et al., 2000).

La levée de répression de la floraison est le résultat de stimuli déclenchant quatre grandes cascades d'activation de la floraison: les cascades vernalisation, photopériode, autonome et GA.

2-La cascade vernalisation

Les plantes ont des besoins différents en vernalisation selon l'espèce, voire même la variété. Nous nous limiterons ici au cas des plantes sensibles à cette exposition au froid, à savoir les plantes bisannuelles et les plantes annuelles d'hiver (Amasino, 2004b).

Avant l'arrivée de l'hiver, le niveau d'expression de l'inhibiteur *FLC* est fort, empêchant ainsi la transition florale (Simpson and Dean, 2002). Puis pendant l'hiver, les plantes doivent non seulement sentir le froid, mais également disposer d'un mécanisme pour mesurer la durée de cette période afin d'éviter de fleurir pendant d'éventuels redoux intempestifs (Sung and Amasino, 2004). *VERNALIZATION INSENSITIVE3* (*VIN3*) est l'acteur central de cette réponse: ce gène codant pour une protéine à homéodomaine de plantes (*PHD finger*) est induit par plusieurs semaines d'exposition à basse température, majoritairement au niveau des méristèmes racinaire et apical (sites de perception du froid; Amasino, 2004a) et son expression s'éteint rapidement après un transfert à température plus clémente. A l'aide de *VERNALIZATION1* (*VRN1*; Levy et al., 2002) et *VERNALIZATION2* (*VRN2*), il réprime de façon durable l'expression de *FLC* en recrutant des histone de-acétylases et méthylases qui modifient la chromatine du locus *FLC* et le stabilise ainsi à l'état réprimé (Fig.3, Sung and Amasino, 2004). Ces modifications épigénétiques serviraient de « marqueurs de vernalisation », assurant la mémoire cellulaire au cours des divisions mitotiques (Amasino, 2004a). A la sortie de l'hiver, le faible niveau d'expression de *FLC* ne peut donc plus s'opposer à l'action des intégrateurs floraux: la floraison peut alors se déclencher (cf paragraphe 1.2.b. 'Les intégrateurs floraux', (Blazquez and Weigel, 2000; Kardailsky et al., 1999; Kobayashi et al., 1999; Lee et al., 2000; Nilsson et al., 1998; Samach and Coupland, 2000).

3-La cascade photopériodique

La cascade photopériodique est sans doute la cascade la plus influente dans le déclenchement de la floraison sous nos régions tempérées. Son étude a fait l'objet de nombreux progrès récents. Elle traduit la capacité de la plante à percevoir si les jours sont longs ou courts.

L'acteur central de cette voie est *CONSTANS* (*CO*). *CO* est régulé par deux processus: i) le niveau d'ARNm de *CO* oscille selon un rythme circadien interne; ii) parallèlement, la lumière stabilise la

protéine CO. Dans le cas des plantes de jours longs (*A. thaliana* notamment), c'est seulement au moment où les ARNm sont présents et les protéines stabilisées, c'est à dire uniquement en fin de journée en photopériode longue, que CO va pouvoir transmettre le signal activateur de la floraison *via* l'induction de l'expression de *FT* (Fig.3, Putterill et al., 1995; Suarez-Lopez et al., 2001; Valverde et al., 2004). Le mécanisme moléculaire de cette activation n'est pas compris. En effet, CO code pour une protéine nucléaire à doigt de zinc B-box qui à elle seule ne semble pas capable de lier l'ADN. Pour agir, des hypothèses suggèrent que CO pourrait intégrer un complexe qui contacte directement le promoteur de *FT* (Ben-Naim et al., 2006; Samach et al., 2000; Wenkel et al., 2006).

Un long mystère a plané quand à la suite de cette voie de transduction: si elle débute au niveau des feuilles, siège des photorécepteurs qui conduisent à la stabilisation de la protéine CO, comment l'information « photopériode favorable à la floraison » circule-t-elle jusqu'au méristème apical où se forment les fleurs? Cette question revient à identifier la nature du célèbre « florigène » (terme introduit par Chailakhyan en 1936). L'hypothèse de l'ARNm CO ou de la protéine CO semble exclue étant donné que l'expression de CO restreinte à l'apex n'a pas d'effet (Corbesier et al., 2007). Des travaux récents ont montré que c'est la protéine FT qui est capable de voyager avec la sève élaborée dans le phloème (Jaeger and Wigge, 2007; Lin et al., 2007; Mathieu et al., 2007) et que cette protéine, une fois dans le méristème apical, active l'expression de régulateurs floraux en s'associant au facteur de transcription FLOWERING LOCUS D (FD; famille des bZIP), gène exprimé seulement au niveau de l'apex (Fig.3, Abe et al., 2005; Wigge et al., 2005).

4-La cascade autonome

La cascade autonome est sous contrôle de signaux endogènes qui évitent à la plante de fleurir à un âge développemental trop faible (Simpson et al., 1999). Elle tire son nom du fait qu'elle agit indépendamment des facteurs environnementaux. La mutation des gènes impliqués dans cette voie de signalisation, *fca*, *fy*, *fpa*, *fve*, *luminidependens (ld)*, *fld*, et *flk*, entraîne un délai dans la floraison indépendamment de la photopériode (Henderson and Dean, 2004). Tout comme la cascade vernalisation, elle a pour finalité la répression de *FLC*. Les acteurs de cette voie agissent par sous groupes épistatiques ([*FCA* et *FY*], [*FVE* et *FPA*], [*LF*], [*FLD*], [*FLK*]) dont certains répriment l'expression de *FLC* par modification post-transcriptionnelle de ses ARNm (Fig.3, Henderson and Dean, 2004; Simpson et al., 1999).

5-La cascade GAs

La promotion de la floraison exercée par les GAs est particulièrement forte en condition de photopériode non-inductive (Wilson et al., 1992). Jusqu'ici il semble que seuls des gènes de biosynthèse des GAs (*GA1*, *GA4*, *GA5*) et des gènes de réponse aux GAs (*GIBBERELLIC ACID INSENSITIVE (GAI)*, *REPRESSOR OF GA1-3 (RGA)* et *RGA-LIKE1 (RGL1)*, *SPINDLY (SPY)*, *PHOTOPERIOD RESPONSIVE1 (PHOR1)*, *FLOWERING PROMOTIVE FACTOR1 (FPF1)*, *SHORT*

INTERNODES (SHI)) participent à cette cascade: aucun acteur aval à cette voie de signalisation agissant spécifiquement dans l'induction florale n'a été identifié. *GAI*, *RGA*, *RGL1* et *SPY* inhibent la réponse GA en absence de GA active. La présence de GA agirait selon des mécanismes non caractérisés impliquant *PHOR1*, *PPF1* et *SHI* pour lever cette inhibition (Fig.3, Mouradov et al., 2002).

Nous venons de survoler les quatre principales cascades d'activation de la floraison. Bien qu'elles constituent l'essentiel du contrôle du temps de floraison, des facteurs secondaires interviennent également, sollicitant le même jeu de régulateurs.

6-Les facteurs secondaires

La chaleur est un accélérateur de floraison, activation qui serait plus influencée par la température ambiante nocturne que diurne (Thingnaes et al., 2003). Ce stimulus agirait *via* la cascade autonome vu que les plantes mutées *fca* et *fve* fleurissent aussi tardivement à 16°C qu'à 23°C. Cette voie met en jeu le répresseur *SVP*, dont la mutation entraîne une insensibilité de la plante aux changements de température. De plus, le fait que *SVP* se lie directement au promoteur de *FT* pourrait expliquer pourquoi la température semble perturber la cascade photopériode (Blazquez et al., 2003; Lee et al., 2007).

Les stress (biotiques, abiotiques) peuvent avoir des effets contraires sur le temps de floraison: certains engendrent une floraison prématurée, réaction de la plante qui lui permettent d'assurer autant que possible le succès de sa reproduction, d'autres la retardent. Nombre de ces stress occasionnent la synthèse de messagers secondaires tels que l'acide salicylique (AS) et l'oxyde nitrique (NO) impliqués dans des réactions de défense. Suite à un traitement UV-C ou une infection pathogène, la floraison est accélérée en conséquence d'une activation de *FT* *via* l'AS. En revanche, une carence en sel provoque un retard de floraison sous l'effet supposé d'une accumulation en NO, agissant sur la régulation positive de *FLC* et la répression de *CO* (Ausin et al., 2005).

Les sucres, le saccharose en particulier, contribuent à accélérer le temps de floraison. Sachant que l'addition de saccharose au niveau des parties aériennes (notamment au niveau de l'apex) annule le phénotype de floraison tardive des mutants *co* et *fca* impliqués respectivement dans les cascades photopériode et autonome, mais n'annule pas celui de *ft*, il semblerait que ce sucre agisse en aval de ces deux cascades mais en amont de *FT* (Bernier and Périlleux, 2005; Roldan et al., 1999). Plusieurs travaux suggèrent également que le saccharose serait nécessaire à l'activation de *LFY* par la cascade GA (Blazquez et al., 1998; Eriksson et al., 2006).

7-Interférences entre cascades

Les cascades d'activation de la floraison exercent des influences les unes sur les autres. Même si les bases moléculaires de ces interférences ne sont pas encore bien comprises, elles trouvent explication dans le fait que beaucoup de gènes de régulation sont communs à différentes cascades. Il

existe par exemple des écotypes d'*A. thaliana* qui germent à la fin de l'été en environnement naturel. Bien que la photopériode soit à ce moment là favorable, *FLC* activé très tôt dans le développement réprime *FT* plus fortement que *CO* ne l'active: la plante ne fleurit pas (Borner et al., 2000; Hepworth et al., 2002; Lee et al., 2000; Searle et al., 2006). Ce n'est qu'à la suite d'une répression de *FLC* via les cascades vernalisation et autonome, c'est à dire au printemps suivant, que la plante pourra être réceptive au signal photopériodique: même si la vernalisation et la cascade autonome en elles-mêmes ne déclenchent pas la floraison, elles constituent des pré-requis indispensables pour rendre la plante apte à fleurir (Amasino, 2004b; Simpson et al., 1999).

b) Les intégrateurs floraux: *LFY*, *FT* et *SOC1*

Les trois intégrateurs floraux *LFY*, *FT* et *SOC1* ont été mentionnés précédemment. Ils sont les éléments terminaux des cascades de floraison et ont pour rôle de centraliser l'information issue de ces voies pour conduire à l'activation des gènes d'identité des méristèmes floraux (Tab.1 et Fig.4; Kobayashi and Weigel, 2007).

1-Présentation et mise en évidence du rôle dans le temps de floraison

La présentation de *LFY*, mon objet d'étude, sera détaillée dans le chapitre que je lui ai consacré (*Chapitre II. 'LFY, un régulateur clé du développement floral'*). Il faudra simplement retenir ici que *LFY* participe à l'induction de la floraison.

Le gène *FT* codant pour le florigène a été isolé simultanément par deux méthodes indépendantes: par l'étude de mutants fleurissant tardivement étiquetés par T-DNA d'une part, et par analyse d'allèles activateurs marqués fleurissement précocément (*activation tagging*) d'autre part (Kardailsky et al., 1999; Kobayashi et al., 1999). Le mutant *ft* fleurit tardivement en jours longs mais semble très peu affecté en jours courts, indiquant un rôle dans la cascade photopériodique. Ce gène code une petite protéine globulaire présentant des homologies de séquence avec la protéine de liaison du phosphatidylethanolamine (PEBP) et la protéine inhibitrice de Raf kinase (RKIP) chez les animaux (Kardailsky et al., 1999; Kobayashi et al., 1999). La fonction de la protéine FT est encore assez mal caractérisée. Il a cependant été montré récemment qu'elle agit comme co-activateur transcriptionnel par liaison directe au facteur de transcription FD (Abe et al., 2005; Wigge et al., 2005).

Le gène *SOC1* code pour un facteur de transcription à boîte MADS. Étonnamment, le mutant *soc1* n'a pas été isolé par criblage génétique classique sur le critère de la floraison tardive, mais par un criblage de suppresseur des effets d'une surexpression de *CO* d'où il tire son nom (*SUPPRESSOR OF OVEREXPRESSION OF CO 1*; Onouchi et al., 2000; Samach et al., 2000). Le mutant *soc1* fleurit néanmoins tardivement indépendamment de la photopériode (Borner et al., 2000; Lee et al., 2000).

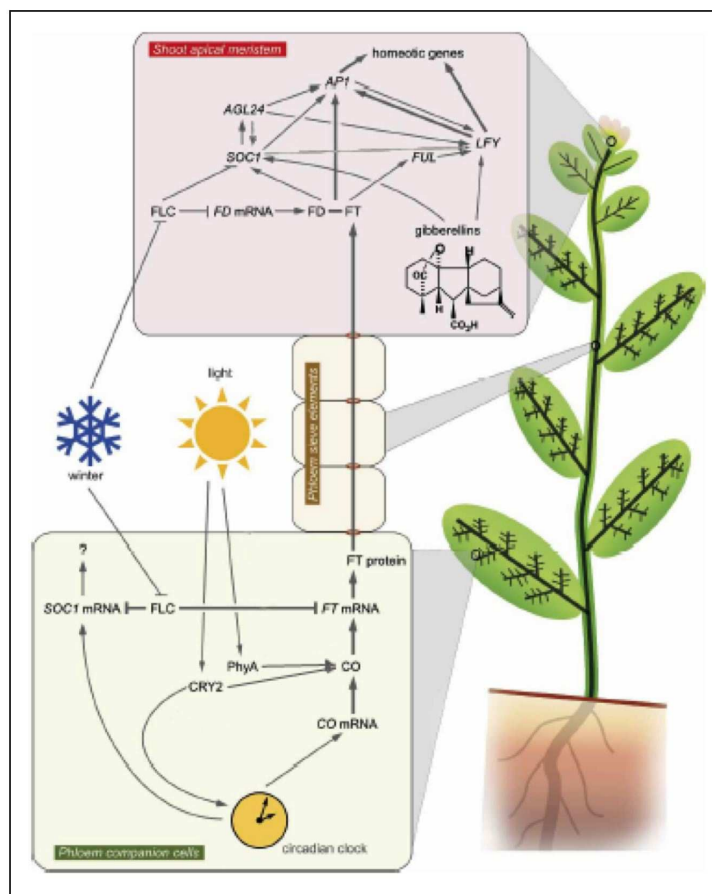


Figure 4: Visualisation spatiale de l'intégration des signaux temps de floraison à l'échelle de la plante (adapté d'après Kobayashi et Weigel, 2007)

Le signal photopériodique est capté par la plante au niveau de ses feuilles. Le signal de la vernalisation agit quant à lui sur FLC, à la fois au niveau des feuilles, et au niveau du méristème apical. Les cascades de signalisation déclenchées par les signaux au niveau de la feuille transitent via le phloème jusqu'au méristème apical. Le signal GAs agit directement au niveau du méristème apical.

Nous allons désormais voir comment les différentes cascades d'activation de la floraison régulent chacun des intégrateurs floraux.

2-Des cascades d'activation aux intégrateurs floraux

-Cascade photopériodique

Nous venons de voir que la cascade photopériode agit essentiellement sur l'expression de *FT*.

Les rôles de *SOC1* et *LFY* dans cette cascade sont en revanche encore très mal définis. Tout comme *FT*, *SOC1* est activé directement par *CO*. Suite à cette activation, *SOC1* s'engagerait dans une voie de transduction indépendante *FT* via *AGAMOUS-LIKE24* (*AGL24*; Liu et al., 2008; Samach et al., 2000; Yu et al., 2004). *SOC1* pourrait également être activé par *FT* (Hepworth et al., 2002; Lee et al., 2000; Samach et al., 2000; Yoo et al., 2005). Quant à *LFY*, il est induit chez les plantes *A. thaliana* exposées à des jours longs mais le mécanisme est peu compris: il serait activé indirectement par *CO* (Blazquez et al., 1997; Nilsson et al., 1998; Simon and Coupland, 1996).

-Cascades autonome et de la vernalisation

Les cascades autonome et de la vernalisation convergent toutes deux vers *FLC* qui réprime *FT*. Cette régulation est en partie directe car *FLC*, facteur de transcription à boîte MADS, contacte le premier intron de *FT*. *FLC* pourrait également agir indirectement *via* un processus dépendant du promoteur même de *FT* (Helliwell et al., 2006; Searle et al., 2006).

FLC joue un rôle répresseur à deux niveaux sur l'expression de *SOC1*: au niveau du méristème apical, il réprime directement *SOC1* par liaison à ses séquences cis-régulatrices (Helliwell et al., 2006; Hepworth et al., 2002; Searle et al., 2006); au niveau de la feuille, il agit *via* une répression de *FT* causant indirectement la répression de *SOC1* (He et al., 2003; Michaels and Amasino, 1999; Searle et al., 2006; Sheldon et al., 1999). *SOC1* pourrait également agir *via* une voie indépendante de *FLC* suite à son activation directe par *AGL24* (Liu et al., 2008; Michaels et al., 2003; Moon et al., 2003; Yu et al., 2002). Il semble donc que *FLC* et *AGL24*, respectivement répresseur et activateur directs de *SOC1*, entrent en compétition pour établir le temps de floraison *via* la régulation de *SOC1* (Liu et al., 2008; Searle et al., 2006).

Jusqu'ici, aucun effet de *FLC* sur *LFY* n'a pu être observé. Cependant, chez les mutants de la cascade autonome, l'expression de *LFY* est réduite et résulterait donc probablement d'une action de *FLC* (Nilsson et al., 1998). D'autre part, *LFY* pourrait intervenir dans la cascade vernalisation *via* la voie de signalisation indépendante de *FLC*: cette voie encore mal caractérisée met en jeu le gène *AGL19* dont l'expression constitutive forte entraîne une surexpression de *LFY* (Schonrock et al., 2006).

-Cascade des GAs

LFY est probablement l'intégrateur floral le plus important pour l'intégration du signal GAs. Les plantes mutées dans le gène de biosynthèse des GAs *ga1-3* ont une expression de *LFY* fortement réduite en jours courts et une surexpression de *LFY* annule leur retard de floraison (Blazquez et al., 1998). La voie des GAs agit sur l'expression de *LFY* *via* un site de réponse aux GAs situé dans le promoteur de *LFY*. Ce site présente des homologies avec un site de liaison des facteurs MYB aux GAs et est reconnu par *AtMYB33 in vitro* (Gocal et al., 2001). La mutation de ce site entraîne l'absence de réponse de *LFY* à une exposition aux jours courts, mais ne perturbe pas sa réponse en jours longs (Blazquez and Weigel, 2000).

Le signal GA induit également l'expression de *SOC1* et *AGL24* qui s'autorégulent directement et agiraient de façon synergique pour activer la floraison en jours courts (Liu et al., 2008; Moon et al., 2003).

Les évidences quant à une implication de *FT* dans cette cascade restent pour leur part fragiles (Gomez-Mena et al., 2001; Pineiro et al., 2003).

Le déclenchement de la floraison résulte également de régulations d'expression entre intégrateurs floraux. En effet, *FT* activerait indirectement *LFY* et *SOC1* (Parcy, 2005; Searle et al., 2006), tandis que *SOC1* active directement *LFY* en se liant à son promoteur (Liu et al., 2008).

Suite à leur induction, les intégrateurs floraux vont pouvoir aller activer à leur tour les gènes d'identité méristématique florale ou FMI (Floral Meristem Identity genes) dont les principaux sont LFY, APETALA1 (AP1), CAULIFLOWER (CAL) et FRUITFULL (FUL).

3-Des intégrateurs floraux aux gènes d'identité méristématique

La régulation la mieux caractérisée est celle d'*AP1*. *AP1* est activé directement par *LFY*, nous y reviendrons plus en détail dans le paragraphe II.2.c. 'Activation de gènes d'identité du méristème floral' (Parcy et al., 1998; Wagner et al., 1999; William et al., 2004). *LFY* n'est cependant pas le seul activateur d'*AP1*: chez le mutant *lfy*, on observe tout de même une expression d'*AP1* mais elle est retardée par rapport à une plante sauvage (Huala and Sussex, 1992; Weigel et al., 1992). L'étude du double mutant *ft lfy* chez lequel *AP1* n'est plus du tout exprimé indique que c'est *FT* qui est responsable de l'induction d'*AP1* indépendamment de *LFY* (Ruiz-Garcia et al., 1997). Effectivement, la capacité de *FT* à activer *AP1* a été mise en évidence par Wigge et collaborateurs (2005) qui montrent que *FT* agit à ces fins en tant que co-facteur du facteur de transcription *FD*. *CAL* est également activé directement par *LFY*, mais aucune activation par *FT* n'a pu être mise en évidence à ce jour malgré des homologies de séquence et une fonction partiellement redondante à *AP1* dans l'identité du méristème floral (Alvarez-Buylla et al., 2006; Kempin et al., 1993; Schmid et al., 2003; Wagner et al., 2004; William et al., 2004).

Enfin, *FT* est le seul intégrateur floral connu pour activer *FUL* (Teper-Bamnolker and Samach, 2005; Tilly et al., 1998)

3. Morphogenèse florale

a) Les gènes d'identité des méristèmes floraux ou gènes FMI

Les gènes FMI, exprimés immédiatement après les intégrateurs floraux, sont responsables du développement des premiers stades des méristèmes floraux dont les principales caractéristiques morphologiques sont décrites en figure 5A. *LFY* fait aussi partie de cette classe mais son cas sera traité en détail plus loin (cf paragraphe II.2.'Rôle dans l'identité méristématique florale'). Nous nous focaliserons ici uniquement sur *AP1*, *CAL*, *FUL*, *UNUSUAL FLORAL ORGANS (UFO)* et *TFL1*.

Les gènes *AP1*, *CAL* et *FUL* codent tous trois pour des facteurs de transcription à boîte MADS de type-II: ils possèdent un domaine M (MADS) en position N-terminale, une région I divergente, un domaine conservé K spécifique des plantes, et une région C (C-terminale) divergente. Le domaine MADS (60 aa) est un domaine de liaison à l'ADN et d'homodimerisation extrêmement conservé

reconnaissant spécifiquement le motif CC(A/T)₆GG, connu sous le nom de « boîte CArG » (Riechmann et al., 1996).

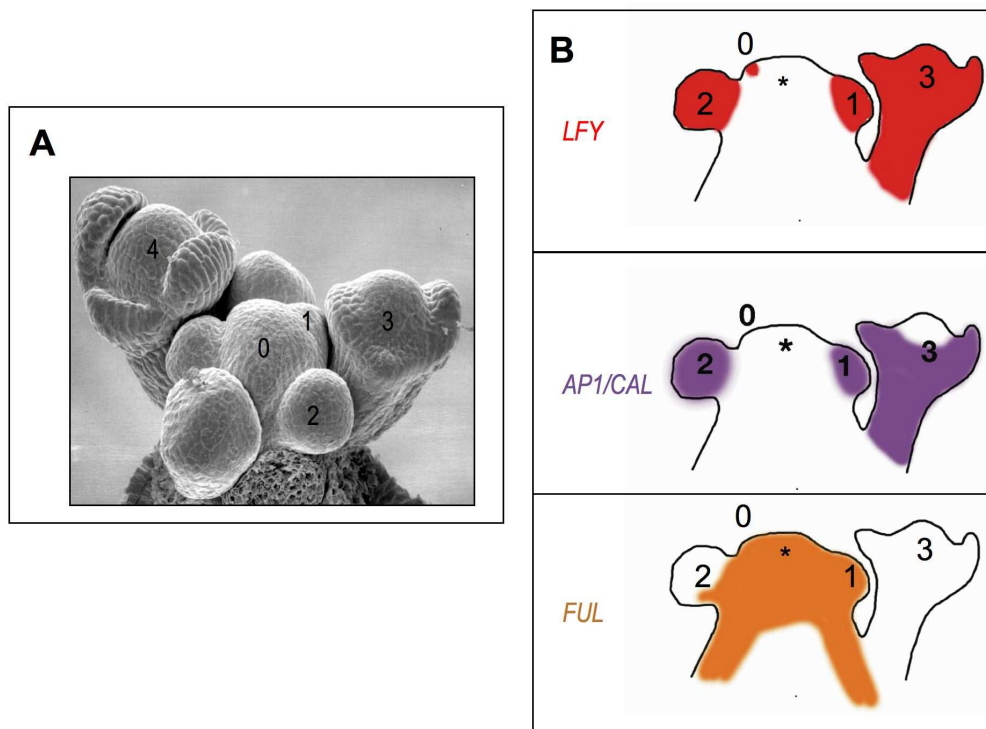


Figure 5: Description des jeunes stades de développement floraux et profil d'expression des principaux gènes FMI.

(A) Photographie en microscopie électronique à balayage d'un méristème d'inflorescence supportant des méristèmes floraux à différents stades de développement (numérotés). L'initiation de la fleur débute lorsque les cellules qui donneront la fleur (stade 0 ou anlagen) forment un contrefort (stade 1) sur le flanc du méristème d'inflorescence nouvellement induit en méristème floral. Suite à la croissance et à la division de ces cellules, une bosse, le primordium méristématique floral, se forme pour se détacher du méristème adjacent (stade 2). Par la suite apparaissent au niveau du primordium méristématique de nouveaux contreforts correspondant aux quatre futurs sépales (stade 3) qui grossissent jusqu'à dessiner des arêtes distinctes (stade 4). Les primordia des quatre pétales et des six étamines s'initient à leur tour au début du stade 5 et se développent au stade 6 parallèlement à l'émergence du futur gynécée; (B) Représentations schématiques des profils d'expression de *LFY*, *AP1/CAL*, *FUL* au niveau des jeunes stades de développement floraux. Les couleurs représentent les ARNm (ou l'activité du promoteur::GUS). L'étoile indique le méristème apical de inflorescence, les chiffres correspondent aux différents stades de développement du méristème floral (d'après Parcy, 2005).

Chez le simple mutant *ap1*, les défauts d'identité méristématique sont assez subtils: l'entre-noeud au niveau des organes floraux est parfois allongé et certaines fleurs sont sous tendues par des bractées (Irish and Sussex, 1990). Le simple mutant *cal* présente quant à lui des fleurs totalement normales (Kempin et al., 1995). En revanche, chez le double mutant *ap1 cal*, le rôle redondant de ces deux gènes dans l'identité des méristèmes floraux est particulièrement net.

Les plantes *ap1 cal* ressemblent fortement (mais en plus petit) au chou-fleur comestible: les inflorescences sont constituées d'une prolifération de jeunes stades floraux qui n'arrivent pas à atteindre le stade de fleurs matures (Kempin et al., 1995). En accord avec ce rôle redondant, *AP1* et *CAL* ont un profil d'expression similaire: tous deux s'expriment très tôt au cours du développement du méristème floral (à partir du stade 1) puis leur expression se restreint à partir du stade 3 aux deux

premières couronnes et au pédicelle de la fleur (Fig.5B, Kempin et al., 1995). Si le mode d'action de *CAL* n'est pas encore bien compris, les données sur *AP1* sont de plus en plus abondantes. D'une part, *AP1* agit en stimulant l'expression de *LFY* pour qu'elle persiste dans la fleur et induit également l'expression des gènes d'identité des organes floraux *APETALA3* (*AP3*) par activation directe et *PISTILLATA* (*PI*). D'autre part, *AP1* réprime au niveau du méristème floral l'expression des gènes d'identité d'inflorescence *TFL1*, *SVP* (répression directe), *AGL24* (répression directe) et *SOC1* (répression directe; Hill et al., 1998; Liljgren et al., 1999; Liu et al., 2007; Yu et al., 2004).

FUL possède une séquence proche de celle d'*AP1* et de *CAL* mais se distingue par son patron d'expression exclu des jeunes stades de développement floraux (Fig.5B). Son rôle en tant que gène FMI n'est apparent que chez le triple mutant *ap1 cal ful*: c'est lui qui permet aux structures en chou-fleur de produire des organes reproducteurs après un temps de prolifération (Ferrandiz et al., 2000). Comme le simple mutant *cal*, le simple mutant *ful* ne présente en revanche aucun phénotype relatif à l'identité du méristème floral (Gu et al., 1998). *FUL* n'apparaît donc pas comme un acteur majeur de ce processus.

UFO code pour une protéine à boîte F, catégorie de protéines connue pour conférer une spécificité de substrat à une classe d'ubiquitine ligases E3 ou SCFs (Wang et al., 2003). Il semble effectivement qu'*UFO* puisse contrôler l'ubiquitination de protéines cibles car il interagit avec une autre sous-unité du complexe SCFs, *ASK1* (Ni et al., 2004; Samach et al., 1999; Zhao et al., 1999). *UFO* intervient principalement dans l'identité des organes floraux, nous y reviendrons dans le paragraphe II.3. *Rôle de FMI tardif dans la morphogenèse florale*. Même s'il est moins important, son rôle dans l'identité méristématique floral a été déduit de plusieurs observations. D'une part, la mutation perte de fonction *ufo* conduit à des changements subtils de la structure d'inflorescence tels l'augmentation du nombre d'inflorescences secondaires et la présence de bractée sous-tendant certaines fleurs. D'autre part, plus démonstrative, la mutation de *ufo* conjuguée à celle de *lfy* ou *ap1* aggrave le phénotype de ces simples mutants (Levin and Meyerowitz, 1995; Wilkinson and Haughn, 1995). Même s'il semble qu'*UFO* agisse en coopération avec *LFY* dans le cas de la suppression des bractées notamment, et qu'il interagisse directement avec *LFY* dans l'identité des organes floraux, le mécanisme moléculaire mis en jeu dans l'identité méristématique n'a toujours pas été établi (Hepworth et al., 2006).

En plus de son rôle déjà évoqué dans la répression de la floraison, *TFL1* joue un rôle opposé à *LFY*, *AP1* et *CAL* dans l'identité méristématique florale. Chez les mutants *tfl1*, les méristèmes d'inflorescence sont convertis en méristème floral: les inflorescences secondaires sont remplacées par des fleurs sous tendues par des bractées, et l'inflorescence primaire est convertie en fleur terminale (Shannon and Meeks-Wagner, 1991). *TFL1* s'oppose donc à l'identité du méristème floral en spécifiant l'identité du méristème d'inflorescence. Il serait un régulateur négatif de *LFY* et *AP1/CAL* puisque chez le mutant *tfl1*, les expressions de *LFY* et *AP1* envahissent le méristème d'inflorescence (Ratcliffe et al.,

1998; Shannon and Meeks-Wagner, 1991). Il est intéressant de noter que FT et TFL1 sont des protéines homologues: leurs rôles opposés pourraient s'expliquer par une compétition pour la régulation d'AP1 (Ahn et al., 2006).

Nous venons de voir que les gènes FMI sont responsables des étapes précoces du développement du méristème floral. Nous allons désormais voir comment se déroule le deuxième processus de prolifération conduisant à l'élaboration de la fleur, processus qui met en jeu les gènes d'identité des organes floraux.

b) Bases du développement de la fleur

Malgré la grande diversité de formes et de couleurs relevées chez les fleurs, toutes présentent une même organisation de base, à savoir quatre organes distincts: les sépales et les pétales constituant le périanthe; les étamines et les carpelles constituant les organes reproducteurs respectivement mâle et femelle. Dans le cas d'*A. thaliana*, ces organes sont organisés en couronnes concentriques, les sépales étant la couronne la plus externe, les carpelles la plus interne.

La fleur naît de l'activité du méristème floral dont le développement est extrêmement stéréotypé, divisé en douze étapes par analyse de la morphogenèse en microscopie électronique à balayage (Fig.5A; Bowman et al., 1989; Smyth et al., 1990).

c) Du modèle ABC au modèle du quartet floral (ou ABCE)

L'identité des organes floraux est déterminée génétiquement par l'expression combinée des gènes d'identité des organes floraux. Ce sont les profils d'expression chevauchants de ces gènes qui constituent les bases du modèle ABC (Bowman et al., 1991; Coen and Meyerowitz, 1991; Pelaz et al., 2000). Les données à l'origine de ce modèle ont été engendrées au début des années 1990 à partir de deux plantes modèles, *Antirrhinum majus*, plus connue sous le nom de gueule de loup ou muflier, et *A. thaliana*. L'universalité de ce modèle a été très largement validée par la suite à partir de résultats chez d'autres espèces angiospermes aussi bien eudicotylédones que monocotylédones (Robles and Pelaz, 2005).

En s'appuyant sur l'analyse phénotypique de mutants homéotiques d'identité des organes floraux, le modèle ABC propose que trois activités, 'A', 'B' et 'C', seules ou en combinaison, spécifient l'identité des quatre couronnes d'organes floraux (Fig.6). L'activité 'A' seule est responsable de l'identité des sépales dans la première couronne, 'A' et 'B' spécifient les pétales dans la seconde, 'B' et 'C' les étamines dans la troisième, et l'activité 'C' seule les carpelles dans la quatrième. Le modèle stipule enfin que 'A' et 'C' aient une fonction antagoniste, et que l'activité 'C' soit responsable du caractère déterminé du méristème floral (Bowman and Meyerowitz, 1991; Coen and Meyerowitz, 1991; Pelaz et al., 2000; Robles and Pelaz, 2005).

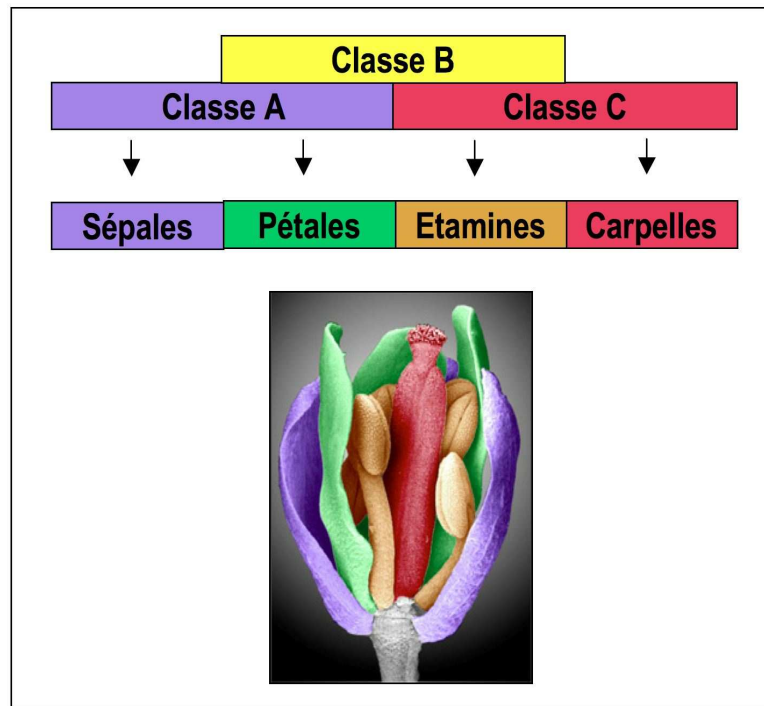


Figure 6 Modèle ABC

Les jeux de couleurs utilisés pour représenter les différentes activités sont conservés entre la représentation du modèle et la photographie d'une fleur d'*A. thaliana* colorée.

Les gènes homéotiques responsables des activités ABC ont été clonés et, constat remarquable, ils appartiennent tous à une exception près à la grande famille des facteurs de transcription à boîte MADS. Tout aussi remarquable, chez les Gymnospermes (plantes à graines non florales telles que les pins), ces protéines à boîte MADS joueraient aussi vraisemblablement un rôle central dans l'identité des organes reproducteurs. Le rôle des gènes 'B' dans l'identité des organes reproducteurs mâles et des gènes 'C' dans l'identité des organes reproducteurs mâles+femelles chez les Angiospermes pourrait être un héritage des Gymnospermes où les orthologues des gènes 'B' et 'C' présentent des profils d'expression en accord avec ces fonctions. Des expériences de complémentation fonctionnelle chez *A. thaliana*, même si elles sont assez rares, vont également dans ce sens (Mouradov et al., 1998; Mouradov et al., 1999; Robles and Pelaz, 2005; Tandre et al., 1995; Tandre et al., 1998; Zhang et al., 2004).

Si nous revenons au cas d'*A. thaliana*, l'activité 'A' est portée par *AP1* et *APETALA2* (*AP2*), ce dernier étant le seul gène du modèle à coder un facteur de transcription spécifique au règne végétal appartenant à la famille de facteur de transcription éponyme. L'activité 'B' est portée par *AP3* et *PI*. L'activité 'C' enfin est portée par *AGAMOUS* (*AG*; Bowman and Meyerowitz, 1991; Coen and Meyerowitz, 1991). Les données génétiques chez *A. majus* obtenues à partir des orthologues *DEFICIENS*(*DEF*)/*GLOBOSA*(*GLO*) respectifs d'*AP3* et *PI* d'*A. thaliana* ont rapidement conduit à l'hypothèse d'une action en dimères en accord avec la capacité des protéines à boîte MADS à se dimériser (cf paragraphe III.1.a.4.' les autres hélices α). Plusieurs hétérodimérisations entre protéines

du modèle ABC ont alors été observées par tests en double hybride chez la levure et par expériences de retard sur gel (ou EMSA, *Electrophoretic Mobility Shift Assay*).

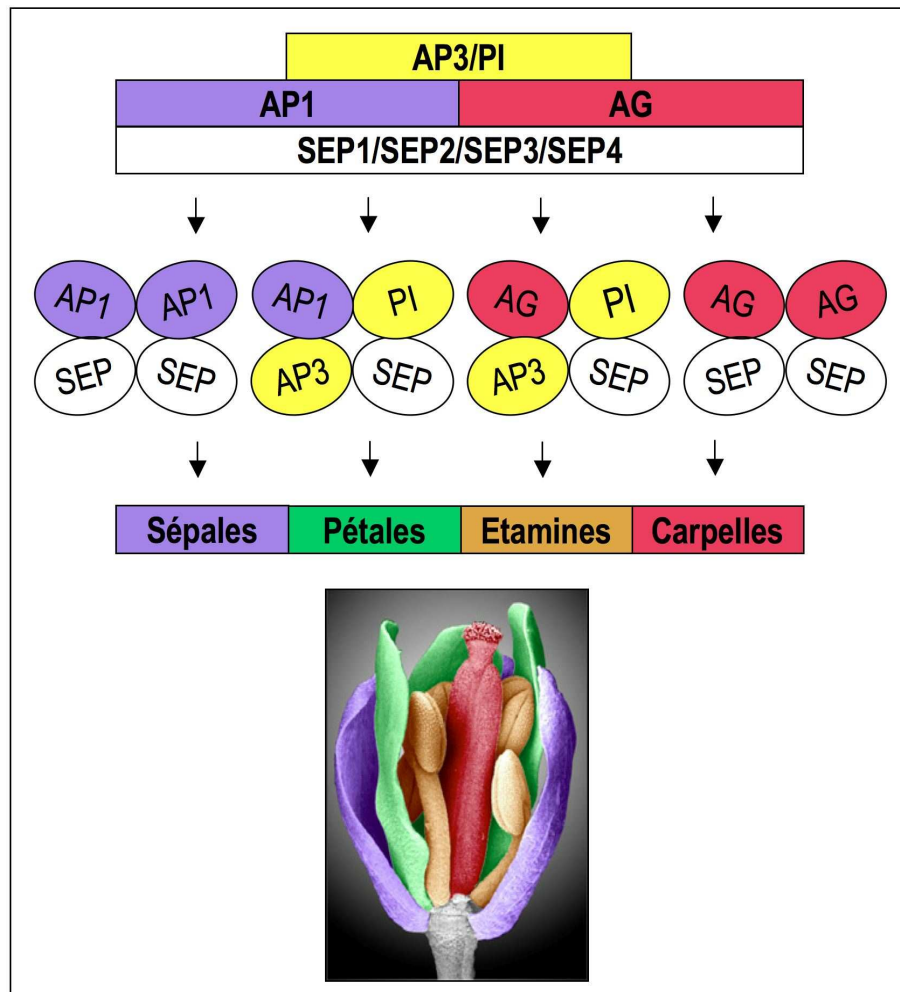


Figure 7: Modèle du quartet floral (sur les bases de Robles et Pelaz, 2005 et Krizek et Fletcher, 2005)

Les jeux de couleurs utilisés pour représenter les différentes activités sont conservés entre la représentation du modèle et la photographie d'une fleur d'*A.thaliana* colorée.

Plus tard, la découverte des 4 gènes à boîte MADS aux rôles redondants *SEPALLATA1* à 4 (*SEP1/4*) est venue compléter le modèle ABC. Si chaque simple mutant *sep* indépendant ne présente aucun phénotype, le quadruple mutant *sep1sep2sep3sep4* n'est plus capable de produire d'organes floraux: la fleur est transformée en structure indéterminée portant des feuilles (Ditta et al., 2004). Le modèle ABC est alors revisité en modèle ABCE où l'activité E, nécessaire à l'identité des 4 organes de la fleur, est portée par les gènes *SEP1/4* (Fig.7; Ditta et al., 2004; Pelaz et al., 2001).

Avec la découverte de cette activité E s'est imposé un nouveau concept: le développement de la fleur résulterait de la formation de larges complexes de facteurs de transcription à l'origine de l'activation de gènes cibles. Le modèle du quartet floral voit le jour (Fig.7). Rapidement, ce modèle va s'imposer suite à des tests réalisés chez la levure qui montrent qu'une protéine MADS s'associe à un dimère. Il est alors postulé qu'AP1/SEP assurent l'identité des sépales, AP1/AP3/PI/SEP l'identité des

pétales, AP3/PI/AG/SEP l'identité des étamines, et AG/SEP l'identité des carpelles (Honma and Goto, 2001; Krizek and Fletcher, 2005; Robles and Pelaz, 2005). Même si ce modèle du quartet floral (Fig.7) est désormais admis par la communauté scientifique, mentionnons que les tétramères n'ont jamais été observés jusqu'ici *in planta*.

Les protéines du modèle ABCE, en tant que facteurs de transcription, contrôlent finalement une cascade de gènes cibles responsables de la différenciation des organes floraux.

II. LFY, un régulateur clé du développement floral

Les rôles de *LFY* à la fois en tant qu'intégrateur floral (décrit dans le paragraphe 1.3.b. *les intégrateurs floraux: LFY, FT et SOC1*) et en tant que gène FMI en font un acteur clé du développement floral. L'expression de *LFY* d'abord détectée dans les jeunes feuilles au cours du développement végétatif augmente graduellement jusqu'à atteindre un seuil suffisant à la transition florale. Cette expression influe donc directement sur le déclenchement de la floraison (Blazquez et al., 1997). Ensuite cette expression perdure dans le méristème floral et *LFY* contribue à bâtir la fleur (rôle FMI).

Le but de ce chapitre est de présenter ce régulateur et faire un bilan des connaissances acquises sur son mode d'action.

1. Présentation

Deux mutants présentant des défauts similaires, à savoir l'absence de fleurs normales et des inflorescences sous-tendues par une feuille à la place des fleurs, ont été décrits chez *A. majus* et *A. thaliana* à la fin des années 1980. *FLORICAULA (FLO)* chez *A. majus* est le premier gène à avoir été caractérisé à partir de l'insertion d'un élément transposable, en 1990 (Coen et al., 1990). L'étude du mutant d'*A. thaliana lfy* obtenu par mutagenèse à l'EMS (éthyl méthanesulfonate) suivra l'année d'après (Schultz and Haughn, 1991), mais ce n'est que deux ans plus tard que l'homologie entre *FLO* et *LFY* sera révélé grâce au clonage du gène chez *A. thaliana* (Weigel et al., 1992). Aujourd'hui, plus de 200 orthologues de *LFY* issus d'espèces différentes ont été clonés.

Dans ce chapitre, je commencerai par évoquer quelques données de base concernant LFY d'A. thaliana: les phénotypes du mutant lfy, les phénotypes de surexpression et le profil d'expression.

Je traiterai ensuite du rôle de LFY en tant que gène FMI précoce (stades floraux 1-2, interaction avec les autres gènes FMI) puis FMI plus tardif (patterning des méristèmes floraux, induction des gènes homéotiques).

a) Description des plantes mutantes *lfy*

Les plantes mutantes *lfy* fleurissent très tardivement, voire jamais. *LEAFY*, littéralement « feuillu », tire son nom du phénotype de ces plantes mutantes où les fleurs sont remplacées par des inflorescences indéterminées soutenues par une feuille. Mais à la différence des mutants *flo* d'*A. majus*, les mutants *lfy*, même les plus sévères, possèdent des structures ressemblant de plus en plus à des fleurs aux positions les plus apicales de l'inflorescence principale (Fig.8). Les allèles mutants *lfy*, 30 identifiés à ce jour, ont ainsi été classés en trois catégories en fonction de la sévérité de leur phénotype (Weigel et al., 1992).

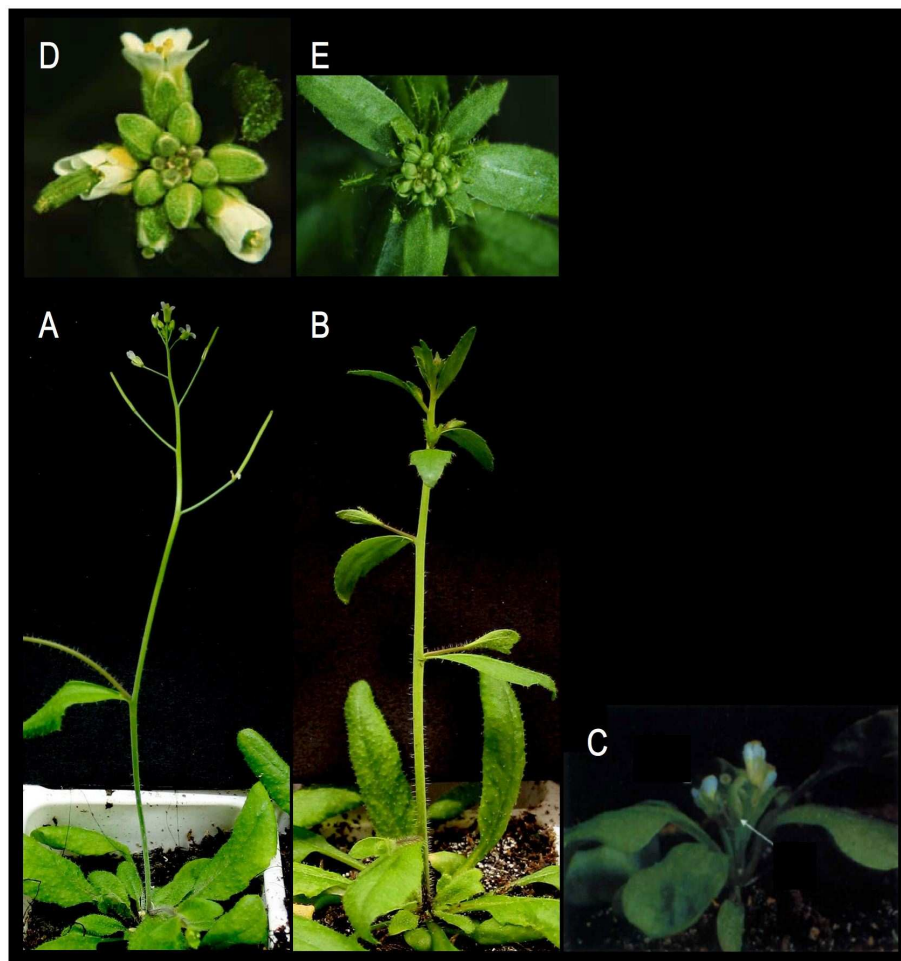


Figure 8: Photographies d'*Arabidopsis thaliana* sauvage (A), *lfy* (B; phénotype fort) et *35S::LFY* (C) et détails des inflorescences sauvages (D) et *lfy* (E; phénotype fort)

Les photographies des plantes sauvage et mutante proviennent d'A. Maizel, celle de la plante *35S::LFY* est tirée de l'article de Weigel et Nilsson, 1995.

1-Phénotype fort

Chez les allèles *lfy* forts, les fleurs les plus basales sont converties en inflorescences secondaires sous-tendues par une feuille caulinaire. Ensuite, en remontant le long de l'inflorescence principale, on trouve des intermédiaires fleurs/inflorescence sous tendues par des « bractées » avec des entrenoeuds allongés et parfois des fleurs qui émergent au sein des fleurs. Enfin, aux positions les

plus apicales, on trouve des structures ressemblant à des fleurs avec une organisation des organes en couronnes, une bractée plus ou moins développées et des organes ressemblant à des sépales ou à une mosaïque sépales/carpelles présentant des trichomes (poils) à trois branches normalement observés sur les feuilles (Fig.8; Weigel et al., 1992).

2-Phénotype intermédiaire et faible

Les fleurs des allèles *lfy* intermédiaires et faibles présentent des phénotypes moins sévères. Les plantes présentent légèrement plus de coflorescences qu'une plante sauvage. Les premières fleurs émergeantes présentent les phénotypes mutants les plus forts comme la présence de fleurs secondaires, mais par la suite les fleurs peuvent développer des pétales et étamines ou organes assimilés. Notons cependant que dans le cas des allèles intermédiaires, pétales et étamines présentent des défauts développementaux résultant en une stérilité mâle. De façon générale, les organes les moins affectés par la mutation *lfy* sont les sépales et carpelles (Weigel et al., 1992).

b) Description des plantes surexprimant constitutivement *LFY* (35S::*LFY*)

Au contraire d'une mutation dans le gène *LFY*, sa surexpression provoque un développement prématuré et ectopique des fleurs. Chez les plantes 35S::*LFY* où l'ADNc de *LFY* est placé sous le contrôle du promoteur fort de l'ARN 35S du virus de la mosaïque du chou-fleur, la tige principale se termine par une fleur et des fleurs solitaires viennent remplacer les tiges secondaires de telle sorte que tout embranchement d'ordre supérieur est exclu. Dans les cas les plus extrêmes, seule une fleur terminale se développe en place de la tige principale immédiatement au niveau de la rosette de feuilles. Ainsi, la surexpression de *LFY* suffit à convertir des tiges en fleurs (Fig.8, Weigel and Nilsson, 1995).

c) Profil d'expression de *LFY*

Le phénotype de surexpression de *LFY* indique que l'expression de ce gène est limitante pour la floraison. Cette hypothèse a été confirmée expérimentalement: l'expression des ARNm *LFY* est faiblement détectée pour la première fois dans les jeunes feuilles au cours du développement végétatif. Cette expression augmente dans chaque nouvelle ébauche de feuilles jusqu'à la transition florale où elle devient maximale. Une fois les fleurs formées, l'expression de *LFY* est détectée de façon uniforme au niveau des jeunes méristèmes floraux (stade 0 ou *anlagen*) jusqu'au stade floral 3 où dès lors que les sépales apparaissent, elle disparaît du centre du méristème floral pour se maintenir au niveau des primordia de ces sépales jusqu'à la fin du stade floral 4. Les ARNm *LFY* sont identifiés en dernier lieu au niveau des primordia des pétales, étamines et carpelles, soit du stade 5 au stade 8 (Blazquez et al., 1997; Coen et al., 1990; Weigel et al., 1992). Précisons que si les ARNm *LFY* disparaissent du centre du méristème floral au cours du stade floral 3, la protéine LFY continue à ce stade à posséder un profil d'expression uniforme qui pourrait résulter d'une stabilité de la protéine et/ou de sa capacité démontrée

à se mouvoir d'une couche cellulaire à l'autre au niveau du méristème (Fig.5B; Parcy et al., 1998; Sessions et al., 2000; Wu et al., 2003).

2. Rôle dans l'identité méristématique florale

Au cours des premiers stades floraux, le rôle des gènes FMI est d'établir l'identité florale au détriment de l'identité de tige/inflorescence. Nous avons vu dans le paragraphe I.4.a) 'Les gènes d'identité des méristèmes floraux ou gènes FMI' qu'il existait d'autres gènes en plus de LFY intervenant dans cette fonction (AP1, CAL, FUL, UFO, TFL1). Lorsque la fonction FMI est compromise, la fleur est partiellement ou entièrement convertie en inflorescences. Nous allons maintenant voir quel est le rôle propre de LFY à ce niveau et comment il interagit avec les autres gènes FMI.

a) Comment agit LFY en qualité de FMI?

Les phénotypes du mutant *lfy* (conversion des fleurs en inflorescences et feuilles) et des plantes qui surexpriment *LFY* (conversion des inflorescences en fleurs) attestent de façon claire du rôle de *LFY* en tant que gène FMI (Coen et al., 1990; Schultz and Haughn, 1991; Weigel and Nilsson, 1995). Les études moléculaires ont montré que *LFY* remplit sa fonction FMI d'une part en interagissant avec d'autres gènes FMI pour assurer la promotion du devenir floral et d'autre part en réprimant des gènes contrôlant le développement des inflorescences au détriment des fleurs.

Nous avons vu précédemment que les gènes *AP1* et *CAL* jouent un rôle clé dans l'identité des méristèmes floraux. Les relations de *LFY* avec *AP1* et *CAL* sont complexes. *LFY* agit déjà en amont de ces deux gènes: son expression augmente graduellement jusqu'à atteindre un seuil suffisant pour activer brutalement *AP1* et *CAL*. Mais *LFY* agit également à leur niveau puisqu'il agit de façon partiellement redondante à *AP1*. *AP1* et *CAL* quant à eux contribuent à assurer un niveau d'expression stable de *LFY* au cours des stades 2-3 du développement des fleurs.

Cette capacité de *LFY* à activer soudainement *AP1/CAL* conjuguée à cette boucle de rétrocontrôle entre *LFY* et *AP1/CAL* sont à l'origine du déclenchement de la transition florale.

b) Une action partiellement redondante à celle d'*AP1*

Toutes les fleurs du mutant *lfy* ne sont pas converties en inflorescences. Aux positions les plus apicales, il existe même chez les allèles forts des fleurs, sans pétales ni étamines, mais possédant des organes sépalloïdes et carpelloïdes disposés selon une phyllotaxie en couronne (Weigel et al., 1992). Ces structures florales disparaissent chez un double mutant *lfy ap1*, indiquant qu'*AP1* est activé et qu'il est capable de conférer une identité florale chez le mutant *lfy*. Des études génétiques ont montré que c'était la voie photopériodique *via* FT et FD qui activait *AP1* en l'absence de *LFY* (Wigge et al., 2005).

La part de cette activation par rapport à celle exercée par *LFY* chez une plante sauvage est difficile à évaluer.

c) Activation de gènes FMI

LFY se situe également en amont d'*AP1* et de *CAL*. Ces deux gènes s'expriment soudainement et uniformément dans le méristème floral à partir du stade floral 1 immédiatement après *LFY* (Kempin et al., 1995; Mandel et al., 1992). Chez le mutant *lfy*, l'expression d'*AP1* est fortement retardée. A l'inverse, chez les plantes 35S::*LFY*, on observe une activation ectopique d'*AP1* dans les feuilles et la mutation *ap1* réduit le phénotype de surexpression de *LFY* en remplaçant les fleurs solitaires par des tiges complexes. *LFY* est donc un régulateur positif d'*AP1* (Liljgren et al., 1999; Weigel and Meyerowitz, 1993b; Weigel and Nilsson, 1995).

En 1998, Parcy et collaborateurs apportent une donnée essentielle quant à la compréhension du mécanisme d'action de *LFY*. Après avoir localisé la protéine au niveau du noyau par immunolocalisation, ils montrent par expérience d'EMSA que *LFY* reconnaît spécifiquement une séquence comprise dans le promoteur d'*AP1*: *LFY* fait donc partie de la grande famille des facteurs de transcription, et régule directement l'expression d'*AP1*. Cette action directe va être confirmée par deux autres approches qui parallèlement montreront également une liaison directe de *LFY* au promoteur de *CAL*. La première repose sur l'utilisation d'une version inductible de la protéine *LFY* fusionnée au domaine de liaison aux hormones du récepteur aux glucocorticoïdes de rat (*LFY-GR*). Suite à un traitement hormonal avec l'hormone synthétique dexaméthasone, ce domaine GR cytoplasmique est transloqué dans le noyau où il permet à *LFY* d'accéder à l'ADN de ses gènes cibles. Après s'être assuré que cette fusion résultait en une forme active de *LFY*, l'ajout de cycloheximide (inhibiteur de la synthèse protéique) a permis de montrer que l'effet de *LFY* sur *AP1* et *CAL* ne nécessitait pas d'étape de traduction intermédiaire (Wagner et al., 1999). La liaison directe de *LFY* à *AP1* et *CAL* a aussi été observée par immunoprécipitation de chromatine (ChIP; William et al., 2004).

Une remarque importante est à faire à ce niveau. En système hétérologue (chez la levure), la protéine *LFY* n'est pas suffisante à elle seule pour activer l'expression d'*AP1* (Parcy et al., 1998). Pour ce faire, *LFY* nécessite d'être fusionnée à un domaine d'activation constitutif (domaine activateur du facteur de transcription VP16 du virus de l'herpès; Triezenberg et al., 1988). De même, chez la plante, l'expression constitutive de *LFY* ne conduit pas à une expression constitutive d'*AP1* (Parcy et al., 1998; Wagner et al., 1999). Il est envisageable que la voie d'induction d'*AP1* dépendant de *FT* puisse jouer le rôle du co-activateur putatif de *LFY* dans l'activation d'*AP1* (Parcy, 2005).

La nécessité d'agir avec un co-régulateur ne semble pas être un cas spécifique à l'activation d'*AP1*. Elle se reproduit dans le cas d'autres activations directes de *LFY*, celle d'*AG* notamment, à laquelle nous reviendrons dans le paragraphe II.3. *Rôle de FMI tardif dans la morphogenèse florale*. De ce constat est né l'idée que *LFY* pourrait agir en tant que facteur de transcription neutre, sans pouvoir

régulateur propre, qui nécessiterait pour réguler l'expression de ses gènes cibles d'agir avec des co-régulateurs (Parcy et al., 1998).

d) Répression de gènes d'identité du méristème d'inflorescence

Nous venons de voir que LFY agit sur l'identité méristématique florale en activant l'expression de gènes impliqués dans cette même fonction, *AP1* et *CAL*. Mais à l'opposé, LFY agit également en réprimant l'expression de gènes d'identité des tiges/inflorescences, à savoir *TFL1* et *AGL24*. Ce mécanisme généralisable aux gènes d'identité méristématique conduit à un modèle assez simple: les gènes d'identité florale répriment l'expression des gènes d'identité des tiges/inflorescences au niveau des méristèmes floraux, et inversement, les gènes d'identité des tiges/inflorescences répriment l'expression des gènes d'identité florale au niveau des méristèmes d'inflorescence (Fig.9; Liljegren et al., 1999; Ratcliffe et al., 1998; Yu et al., 2004).

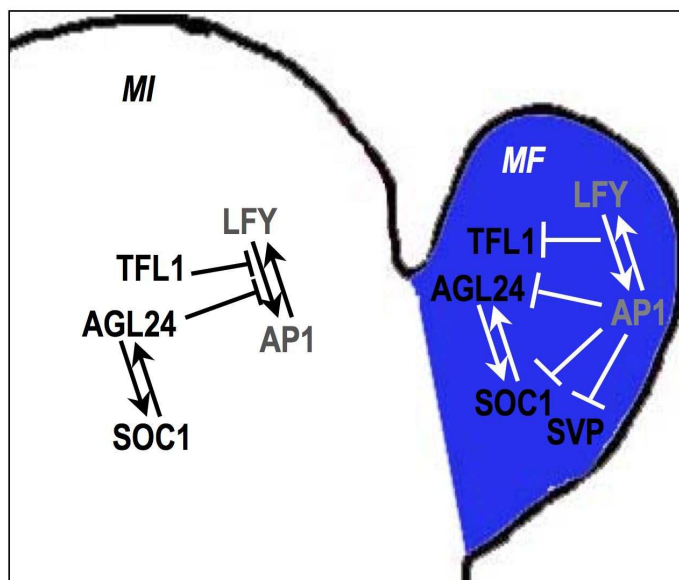


Figure 9: Modèle des réseaux de gènes régulant la mise en place des méristèmes d'inflorescence et floraux

MI = méristème d'inflorescence (zone blanche); MF = méristème floral (zone bleue). Les gènes d'identité des méristèmes d'inflorescence sont représentés en noir, les gènes d'identité des méristèmes floraux sont représentés en gris.

Chez un mutant *lfy*, *TFL1* et *AGL24* sont exprimés ectopiquement au niveau de structures intermédiaires tiges/fleurs qui remplacent les fleurs sauvages (Liu et al., 2007; Ratcliffe et al., 1999), ce qui indique que LFY réprime normalement ces gènes chez une plante sauvage. Aucune interaction directe entre LFY et les séquences cis-régulatrices de ces gènes n'a pu cependant être mise en évidence à ce jour.

Dans la mesure où *AGL24* est réprimé directement par *AP1*, il est très possible que LFY agisse indirectement via l'activation d'*AP1* (Liu et al., 2007).

Dans le cas de *TFL1*, la version activée de LFY (LFY:VP16) active *TFL1* au lieu de le réprimer. A la place des quatre sépales au niveau de la première couronne, les plantes *LFY:VP16* produisent un

nombre anormal d'organes, souvent supérieur à quatre. Les primordia des pétales et étamines sont très mal délimitées et tendent à être remplacées par de nouveaux primordia d'organes (carpelles anormaux essentiellement) adoptant une phyllotaxie légèrement spiralée. Ces observations d'une prolifération partiellement indéterminée du méristème floral et d'une phyllotaxie spiralée soulignent une transformation partielle du méristème floral en méristème d'inflorescence (Parcy et al., 2002). Ces défauts phénotypiques corrélerent avec une expression ectopique de *TFL1* au niveau des primordia floraux des plantes *LFY:VP16*, ce qui suggère que chez une plante sauvage, *TFL1* est directement réprimé par *LFY*.

Nous avons vu que LFY était exprimé au niveau du stade anlagen floral où il joue son rôle de gène FMI précoce et probablement une grande partie de son rôle d'intégrateur floral, ces deux fonctions étant étroitement liées. Mais nous allons voir que LFY, aussi exprimé au niveau des jeunes stades de développement du méristème floral, agit également en tant que gène FMI tardif en activant l'expression des gènes homéotiques déterminant l'identité des organes floraux.

3. Rôle de FMI tardif dans la morphogenèse florale

a) Mise en évidence

L'hypothèse d'une fonction de *LFY* dans l'identité des organes floraux est née en 1990 de la détection par hybridation *in situ* de *FLO* d'*A. majus* au niveau des primordia des sépales, pétales et carpelles (Coen et al., 1990). Huala et Sussex (1992) réitérèrent cette hypothèse chez *A. thaliana* en constatant que les structures florales qui parviennent à se développer chez les mutants fort *lfy* présentent des anomalies d'identité des organes floraux, à savoir l'absence totale de pétales et étamines. A cette même période, l'étude phénotypique de plantes mutées à la fois dans *lfy* et dans un gène d'identité des organes floraux (*ap2*, *ap3*, *pi* et *ag*), montrent que *lfy* est tout ou partie épistatique sur les mutations *ap2*, *ap3*, *pi* et *ag* en fonction du phénotype étudié: *LFY* agit donc en amont de ces gènes en accord avec ce rôle de FMI tardif (Coen et al., 1990; Huala and Sussex, 1992; Schultz and Haughn, 1991; Weigel et al., 1992).

Je m'attacherai tout particulièrement ici à expliquer le profil d'expression des gènes ABC qui, confinés à des territoires spécifiques du méristème floral, tient une grande importance dans la délimitation de chacun des organes floraux (Fig.10).

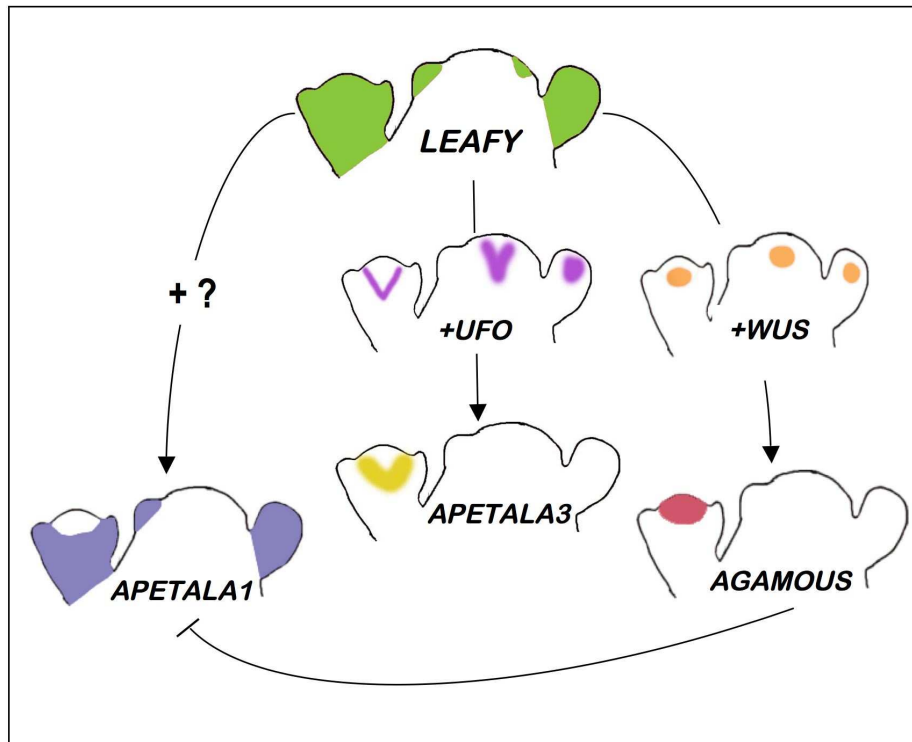


Figure 10: Modèle de délimitation du profil d'expression des gènes d'identité des organes floraux
Représentations schématisques des profils d'expression de *LFY*, *AP1*, *AP3*, *AG*, *UFO* et *WUS* au niveau des jeunes stades de développement floraux. Les couleurs représentent les ARNm (ou l'activité du promoteur::GUS).

b) Rôle de LFY dans l'activation d'*AP1*, gène 'A' d'identité des sépales et pétales

Je ne reviendrai pas ici à l'activation directe exercée par LFY sur le gène 'A' *AP1* qui en plus de son rôle de FMI est impliqué dans l'identité des sépales et des pétales (*cf paragraphes II.2.c. 'Activation de gènes FMI'*). Le profil d'expression d'*AP1*, uniforme dans le méristème floral à partir du stade 1 puis restreint aux primordia des sépales et pétales au stade 3, s'explique par une régulation spatiale exercée d'une part par cette activation de LFY et d'autre part par une répression par AG (Fig.10, Mandel et al., 1992).

c) Rôle dans l'activation d'*AP3* et *PI*, gènes 'B' d'identité des pétales et étamines

Les gènes 'B' *AP3* et *PI* sont exprimés dans les couronnes 2 (pétales) et 3 (étamines) à partir du stade 3 (Mandel et al., 1992). Chez les mutants *lfy*, les pétales et les étamines sont les organes les plus affectés, voire totalement absents chez les allèles forts. Par hybridation *in situ*, ce phénotype se traduit par une réduction du niveau et du domaine d'expression d'*AP3* et *PI* (Weigel et al., 1992; Weigel and Meyerowitz, 1993a).

La compréhension du mécanisme adopté par LFY pour réguler *AP3* a été difficile et laisse subsister beaucoup d'interrogations. En effet, ni la surexpression (35S::*LFY*) ni l'expression de la version de LFY fusionnée au domaine activateur VP16 (35S::*LFY:VP16*) n'occasionne de modification du profil d'expression d'*AP3*. Un précieux indice a été obtenu par la découverte d'un mutant d'*A. majus*,

fimbriata (*fim*), et d'un mutant d'*A. thaliana*, *ufo*, dont le phénotype révèle des défauts de développement des étamines et des carpelles parallèlement à une réduction de l'expression des gènes B (Ingram et al., 1995; Lee et al., 1997a; Simon et al., 1994). Par ailleurs, l'expression de *UFO* est détectée à partir du stade floral 2, au moment où l'identité des pétales et étamines se détermine (Laufs et al., 2003).

Les études de surexpression simultanée de *LFY* et *UFO* ont montré que *UFO* était le corégulateur requis par *LFY* pour activer *AP3*. Le patron d'expression d'*AP3* résulte ainsi de la co-expression de *LFY* et *UFO* au niveau des jeunes méristèmes floraux (Honma and Goto, 2000; Lamb et al., 2002; Lee et al., 1997a; Parcy et al., 1998). Le mécanisme mis en jeu n'est néanmoins pas encore bien compris. Le fait qu'*UFO* soit potentiellement impliqué dans un processus de dégradation des protéines via un complexe d'ubiquitination laisse à penser que *UFO* puisse cibler un répresseur d'*AP3* et *PI*. Un tel régulateur n'a cependant jamais été identifié. Récemment, il a été montré que *LFY* et *UFO* interagissent physiquement et que *LFY* serait ubiquitinylée par des ligases E3 dont *UFO* (Chae et al., 2008). Cette donnée est particulièrement intéressante car elle pourrait s'intégrer aux descriptifs récents répertoriant des cas d'activation de facteur de transcription par ubiquitination selon deux mécanismes distincts: i) chez certains régulateurs, l'ubiquitination conduit à leur dégradation et ainsi à une accélération de leur *turn over* (Kodadek and Bachhawat-Sikder, 2006); ii) chez d'autres facteurs, une monoubiquitination entraîne directement leur activation (Garcia-Higuera et al., 2001).

Par ailleurs, bien que *LFY* semble donc activer directement *AP3*, il pourrait également agir indirectement via *AP1*, autre acteur de l'activation d'*AP3* (Ng and Yanofsky, 2001).

d) Rôle dans l'activation d'AG, gènes 'C' d'identité des étamines et carpelles

L'expression d'AG est détectée à partir du stade 3, délimitée aux couronnes 3 (étamines) et 4 (carpelles). Chez les mutants *lfy*, l'expression précoce d'AG dans le méristème floral est retardée et confinée à un domaine plus petit que celui d'une fleur sauvage (Weigel and Meyerowitz, 1993a).

Un grand pas dans la compréhension du mécanisme d'activation d'AG exercé par *LFY* a été fait à partir de l'étude de plantes exprimant une version activée de *LFY* par fusion au domaine activateur de la protéine virale VP16. Chez les plantes *LFY:VP16* en effet, l'essentiel du phénotype observé résulte d'une suractivation précoce et ectopique d'AG, se manifestant par la formation de structures carpelloïdes ectopiques (Parcy et al., 1998). Alors que la surexpression de *LFY* (*35S::LFY*) ne modifie pas le profil d'AG, la surexpression de *LFY:VP16* suffit à induire AG dans l'ensemble d'une très jeune plantule. Une telle efficacité du domaine VP16-AD suggérerait que *LFY* ait besoin d'un coactivateur. La protéine à homéodomaine *WUSCHEL* (*WUS*), initialement décrite pour son rôle dans la maintenance des cellules souches méristématiques, sera identifiée comme ce coéregulateur trois années plus tard à partir de son profil d'expression au centre du méristème floral d'une part, et du phénotype des plantes

mutantes *wus* très affecté dans le développement des étamines et carpelles d'autre part (Laux et al., 1996; Lohmann et al., 2001; Mayer et al., 1998).

LFY et WUS activent directement et selon un processus synergique l'expression d'AG (Lohmann et al., 2001). Cependant, le fait que le patron d'expression d'AG ne reflète pas exactement celui de WUS et que l'expression d'AG soit toujours détectée chez les mutants *wus* suggère que WUS n'est pas l'unique corégulateur impliqué (Lenhard et al., 2001).

e) Modèle de délimitation du profil d'expression des gènes d'identité des organes floraux

Aujourd'hui, il est acquis que LFY active directement *AP1*, *AP3* et *AG* à l'aide de co-régulateurs conférant une délimitation spatiale du profil d'expression de ces gènes cibles (Honma and Goto, 2000; Lamb et al., 2002; Laufs et al., 2003; Lohmann et al., 2001; Parcy et al., 1998). Ces découvertes ont conduit à l'élaboration d'un modèle génétique dans lequel LFY joue un rôle central au développement des organes de la fleur en activant directement les gènes ABC (Fig.10).

Absents de ce modèle, des données de génomique suggèrent que *LFY* activerait également les gènes *SEP* (Schmid et al., 2003; Wagner et al., 2004). Les données ont aussi permis d'identifier de nouvelles cibles directes de LFY dont le rôle potentiel dans la morphogenèse florale, devrait alimenter le travail de ces prochaines années (Wagner et al., 2004; William et al., 2004).

Nous allons désormais voir dans quelle mesure ces fonctions de LFY établies à partir de la plante modèle A. thaliana sont conservées chez d'autres espèces angiospermes.

4. Un rôle clé généralisable aux Angiospermes

Les études relatives aux orthologues de *LFY* chez de nombreuses espèces angiospermes soulignent que la fonction de LFY en tant que régulateur clé du développement floral semble valable chez la plupart des Angiospermes. Nombre de ces études se restreignent à cloner l'ADNc et dans certains cas à analyser le profil d'expression de ces orthologues, je ne m'y attarderai pas, bien qu'elles soient en accord avec une possible conservation fonctionnelle, pour ne retenir que les arguments directs.

Les meilleurs arguments du rôle des orthologues de *LFY/FLO* dans le développement floral proviennent des espèces où des mutants dans le ou les gènes *LFY/FLO* sont disponibles. En plus des Brassicacées ('*lfy*' d'*A. thaliana*) et des Scrophulariacées ('*flo*' d'*A. majus*), ces mutants existent chez les Fabacées ('*unifoliata*' de pois, '*single leaflet1*' de luzerne, '*LjLFY*' de lotus), les Solanacées ('*aberrant leaf and flower*' de petunia, '*falsiflora*' de tomate) et les Poacées ('*zfl1*' et '*zfl2*' de maïs). Tous présentent au moins des défauts d'identité du méristème floral et des organes floraux (pétales et étamines en particulier), et pour certains un retard de floraison (Bombliès et al., 2003; Dong et al., 2005; Hofer et al., 1997; Molinero-Rosales et al., 1999; Souer et al., 1998; Wang et al., 2008).

D'autres stratégies ont été utilisées, je ne proposerai ici qu'une exposition de quelques cas qui ne constituent en rien une liste exhaustive. L'ADNc d'orthologues de Salicacées, Myrtacées et Rosacées ont été introduits ectopiquement chez *A. thaliana* sauvage et engendrent une floraison précoce de la plante ('*PTLF*' de peuplier, '*ELF1*' d'Eucalyptus; '*AFL2*' et dans une moindre mesure '*AFL1*' de pommier (Rottmann et al., 2000; Southerton et al., 1998; Wada et al., 2002).

Cette même expérience réalisée cette fois chez des *A. thaliana* mutante *lfy* révèle des complémentations fonctionnelles partielles à totales à partir d'orthologues d'Euphorbiacées, Pinacées, Brassicacées et Fabacées ('*HbLFY*' d'Hévéa, '*CfLFY*' du cèdre tropical, '*RFL*' de riz, '*LcrLFY*' de moutarde *Leavenworthia crassa* et '*SGL1*' de luzerne (Chujo et al., 2003; Dornelas and Rodriguez, 2005; Dornelas and Rodriguez, 2006; Sliwinski et al., 2006; Wang et al., 2008).

Soulignons cependant que tous ces orthologues *LFY* n'agiraient pas nécessairement par les mêmes mécanismes. Par exemple, à la différence de *LFY*, *RFL* (*LFY* de riz) est capable d'activer ectopiquement l'expression d'*AP1* et *AG* dans les tissus végétatifs (Chujo et al., 2003). Ou encore, la surexpression de *PTLF* (*LFY* de peuplier) n'entraîne pas de floraison précoce chez le peuplier (Rottmann et al., 2000).

Nous avons vu que LFY jouait un rôle clé dans la régulation de la floraison. J'ai mentionné que la protéine LFY était un facteur de transcription contactant directement les séquences cis-régulatrices de ses gènes cibles. Avant de plonger plus en profondeur dans le fonctionnement moléculaire de cette protéine LFY, je vais donner quelques éléments de description des facteurs de transcription.

III. LFY, un facteur de transcription original

1. Généralités sur les facteurs de transcription

Nous avons vu combien la régulation spatio-temporelle de l'expression des gènes pouvait être importante dans les voies de signalisation de réponses à des stimuli environnementaux ou endogènes, combien elle était importante au développement de la plante. Ce rôle des facteurs de transcription semble généralisable à tout organisme vivant: les facteurs de transcription sont les « architectes du Vivant ». Aussi, l'altération de l'expression des gènes qui les codent serait la source principale de diversité et changement qui sous-tendent l'évolution (Carroll, 2000).

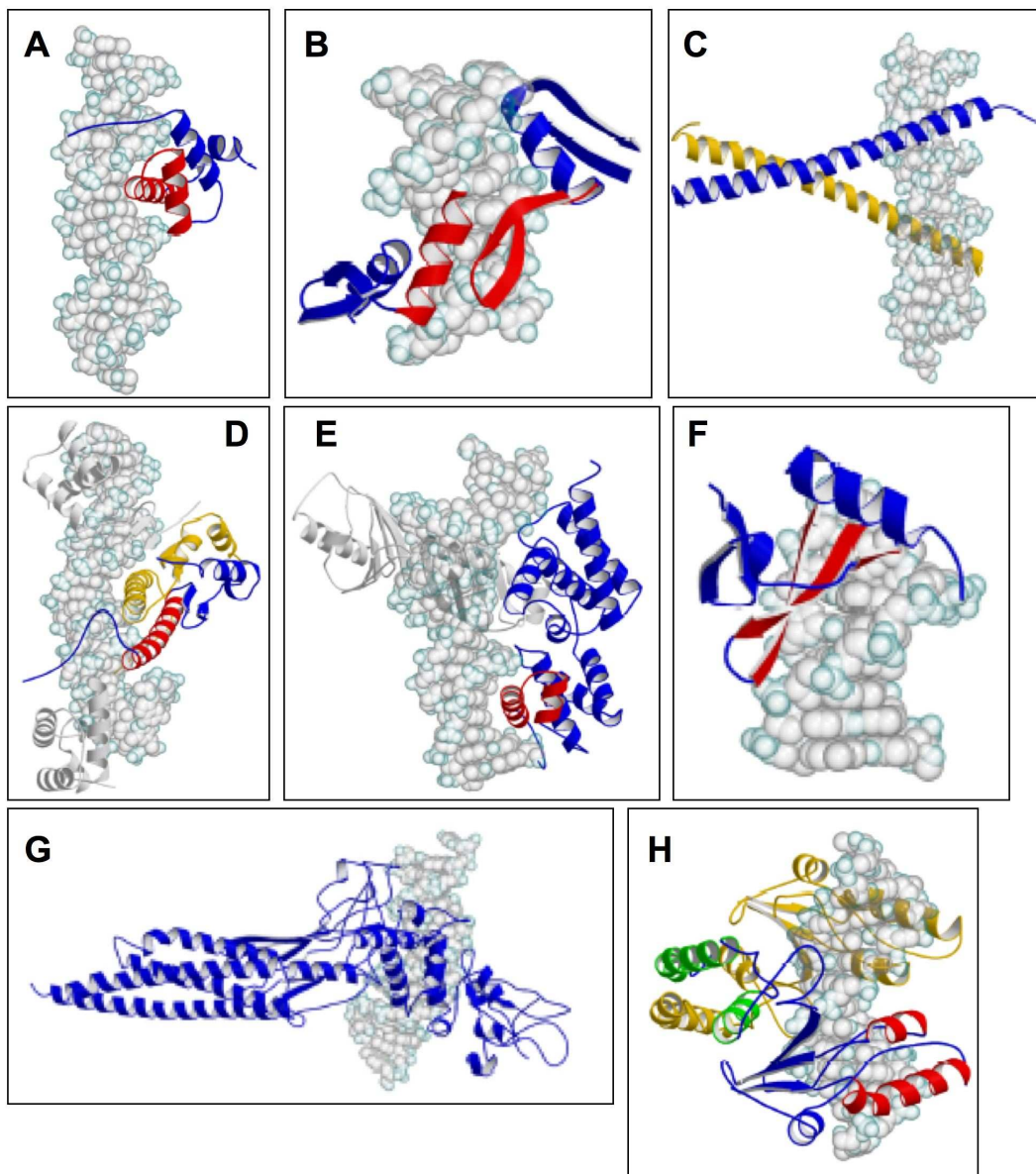


Figure 11: structures des DBD des grandes familles de facteurs de transcription (selon Luscombe et al., 2000)

Dans chaque cas le DBD stricte est indiqué en rouge, le premier monomère en bleu, le deuxième en jaune. La double hélice d'ADN est représentée en gris. (A) Ex. de DBD HTH, l'homéodomaine; (B) Ex. de DBD coordinateur de zinc, le motif doigt de zinc $\beta\beta\alpha$; (C) Ex. de protéine type zipper, une protéine leucine-zipper; (D) Ex. de DBD à autres hélices α , cas des boîtes MADS; (E) Ex. de protéine à feuillet β , une protéine de liaison des boîtes TATA; (F) Ex. de protéines à épingle à cheveux/ruban β , le DBD d'une protéine hyperthermophile chromosomale; (G) Ex. de DBD 'autre', cas des protéines Stat; (H) Ex. de DBD d'enzymes, cas de la famille des endonucléase *PvuII*.

a) Les grandes familles de protéines associées à l'ADN

Les protéines capables de lier l'ADN, à partir des entrées de 240 structures de complexe protéine/ADN dans la banque PDB (*Protein Data Bank*), ont été classés en 8 grandes familles sur les bases de la structure de leur motif de liaison à l'ADN ou DBD (DBD, *DNA binding domain*): 1-les HTH (Hélice-Tour-Hélice, incluant les HTH 'winged', littéralement 'aillés'), 2-les coordinateurs de zinc, 3-les protéines type zipper, 4-les autres hélices α , 5-les feuillet β , 6-les épingle à cheveux/ruban β , 7-les autres, 8-les enzymes (Fig.11; Luscombe et al., 2000).

1-les HTH

La grande famille des HTH naît d'un DBD de 20 aa définissant deux hélices α quasi perpendiculaires connectées par un tour de quatre résidus β . Parmi les HTH sont notamment répertoriées les protéines à homéodomaine, les protéines MYB, la famille des Cro et répresseurs, la famille des répresseurs LacI, les protéines à domaine paire Prd, la famille de transposases Tc3. Par extension, les HTH ailés, présentant un lien du type boucle formée d'une troisième hélice α et d'un feuillet β adjacent qui vient remplacer le tour de résidus β , peuvent être assimilés à des membres de cette famille dès lors que les deux hélices α restent quasi-perpendiculaires. Ce motif contacte l'ADN au niveau du grand sillon *via* la seconde hélice α du motif, élément de reconnaissance ou sonde de l'ADN, qui s'insère véritablement dans le sillon. Les contacts de soutien au squelette phosphate de l'ADN sont principalement assurés par le tour et par la première hélice α qui pont le grand sillon d'ADN à partir de son extrémité N-terminale. D'autres contacts avec l'ADN peuvent être réalisés par le reste de la protéine, et peuvent contribuer également à la spécificité de reconnaissance. Les deux hélices α du motif HTH sont typiquement confinées au sein d'un fagot de trois à six hélices α qui confèrent un coeur hydrophobe stabilisateur. Bien que les motifs HTH puissent présenter de très fortes homologies de structures entre protéines de familles fonctionnelles éloignées, les séquences d'acides aminés sont quant à elles très variables et expliquent la variabilité de séquences d'ADN reconnues (Fig.11A).

En général, il a été observé que les HTH d'origine procaryote reconnaissent des séquences palindromiques sous forme d'homodimère, tandis que les HTH eucaryotes, dont la famille des protéines à homéodomaine, contactent des séquences d'ADN pouvant être non symétriques sous forme soit de monomère, homo- ou hétérodimère.

2-les coordinateurs de zinc

Les protéines coordinatrices de zinc contiennent la plus importante famille de facteurs de transcription eucaryotes, la famille des doigts de zinc $\beta\beta\alpha$. On y compte aussi la famille des récepteurs hormonaux (hormones stéroïdes, thyroïdes et rétinoïdes), la famille boucle-feuillet-hélice (ex.p₅₃) et la famille Gal4. Leur DBD est caractérisé par la coordination tétraédrique de un ou deux atome(s) de zinc par des résidus cystéines et histidines conservés. L'étendue de cet arrangement serait due à la stabilité structurale que confère le ou les atome(s) de zinc. Il est important de noter que la coordination d'atomes de zinc n'est pas seulement observée dans l'interaction protéine-ADN mais également dans des interactions protéine-protéine. La liaison à l'ADN se fait de façon générale par insertion d'une hélice α dans le grand sillon. Une seconde hélice α peut éventuellement venir contacter le squelette phosphate (cas des récepteurs hormonaux et de la famille Gal4) et des boucles peuvent établir des liaisons au niveau du petit sillon (cas de la famille boucle-feuillet-hélice; Fig.11B).

Enfin, ces facteurs peuvent agir en monomère (cas de la famille des doigts de zinc $\beta\beta\alpha$), en homodimère (cas des récepteurs hormonaux et de la famille Gal4), en hétérodimère (cas des récepteurs hormonaux) ou encore en tétramère (cas de la famille boucle-feuillet-hélice).

3-les protéines de type-zipper

Ce groupe tire son nom de la façon dont ses membres se dimérisent. Il regroupe deux familles de protéines d'origine eucaryotes, les *leucine zippers* et les hélice-boucle-hélice ou HLH (*Helix-Loop-Helix*) qui peuvent agir aussi bien en homo- qu'en hétérodimère.

Les *leucine zippers*, les protéines C-Jun et C-Fos par exemple, ont une structure en deux domaines: un domaine de dimérisation et un DBD. Chaque monomère consiste en une grande hélice α d'environ 60 aa. La région *zipper*, ou région *coiled coil*, est constituée de leucines ou de résidus hydrophobes similaires espacés tous les huit résidus afin de toujours être exposés sur la même face de l'hélice α . Ces résidus peuvent ainsi contacter les résidus hydrophobes équivalents au niveau du deuxième monomère. Le domaine DBD, domaine basique situé à l'extrémité N-terminale de chacun des deux monomères, contacte le grand sillon de l'ADN (Fig.11C).

Les HLH dérivent des *leucine zipper* par l'intercalation d'une boucle au niveau de la grande hélice α qui sépare physiquement le domaine de dimérisation du DBD. Hormis cette différence qui leur confère un gain de flexibilité dans le positionnement respectif de chaque hélice N-terminale au niveau de l'ADN, leurs caractéristiques sont les mêmes que celles des leucine-zippers.

4-les autres hélices α

Ce groupe compte sept familles aux fonctions très différentes: Skn-1 et les protéines à boîte MADS sont des régulateurs d'origine eucaryote, Papillomavirus1 E2 et EBNA1 sont des régulateurs et initiateurs de la réplication d'origine virale, les histones et le groupe de protéines à haute mobilité (HMG, *High-Mobility Group*) sont des protéines architecturales de compaction de l'ADN, et Cre est une recombinase site-spécifique. Bien que leurs structures soient très différentes, toutes ces protéines contactent l'ADN via une hélice α . Du fait de leur variabilité, je me limiterai ici au descriptif de protéines présentant un intérêt dans le cadre de mon travail.

Les protéines à boîte MADS, protéines au rôle déjà évoqué à maintes reprises dans les paragraphes précédents, ont un DBD de 56 résidus projetant de longues hélices dans le grand sillon de l'ADN à la manière des *leucine zippers* et établissent d'autres points de contacts avec le squelette phosphate (Fig.11D).

La recombinase Cre est dimérique. Chaque monomère consiste en deux domaines structuraux qui se replient selon un fagot d'hélices. Ensemble ces domaines viennent clamper l'ADN, insérant des hélices α à la fois dans les petit et grand sillons.

5-les protéines à feuillets β

Ce groupe se limite à la famille des protéines de liaison des boîtes TATA. Il se caractérise par un large feuillet β de 10 brins antiparallèles qui vient recouvrir l'ADN au niveau du petit sillon (Fig.11E).

6-les protéines à épingles à cheveux/ruban β

Ce groupe se distingue des protéines à feuillets β par la taille du feuillet ou épingle à cheveux β qui contacte l'ADN, de deux à trois brins. L'interaction peut se faire ici aussi bien au niveau du petit que du grand sillon, en monomère ou en dimère selon la protéine (Fig.11F).

7-les autres

Deux familles non-enzymatiques utilisent un motif structural secondaire de liaison à l'ADN mal défini. Toutes deux fonctionnent en dimère où chaque monomère présente un multidomaine servant à la fois à la liaison à l'ADN, à la dimérisation, et à l'adressage au noyau (Fig.11G).

8-les enzymes

A la différence des autres groupes, ce groupe rassemble des protéines de même fonction (toutes altèrent la structure de l'ADN par catalyse chimique). Chez les enzymes comme chez le groupe des « autres », le DBD est généralement difficile à décrire en termes de motif structural simple, ces protéines utilisant une large combinaison d'hélices α , feuillets β et boucles à ces fins. Beaucoup possèdent trois domaines distincts: un domaine de reconnaissance de l'ADN qui « lit » la séquence d'ADN, un domaine catalytique comprenant le site actif de l'enzyme et si nécessaire, un site de dimérisation. La structure en résultant présente une cavité en forme U où est lié l'ADN, occasionnant souvent de sévères déformations de la structure d'ADN (Fig.11H).

b) Etat des connaissances sur les facteurs de transcription végétaux

Le séquençage du génome entier d'*A. thaliana* a permis d'évaluer la proportion des facteurs de transcription à près de 6% des gènes codants d'*A. thaliana*, soit 1500 gènes environ sans compter les gènes auquel il n'a pas encore été assigné de fonction (40 à 50% des gènes). Cette proportion est 1.3 fois plus importante que celle de la drosophile *Drosophila melanogaster*, et 1.7 fois plus que celle de *Caenorhabditis elegans* et de la levure *Saccharomyces cerevisiae*. De plus, on note que parmi ces 1500 gènes, près de la moitié appartiendrait à des familles de facteurs de transcription spécifiques au règne végétal, innovation qui résulterait du mode de vie végétal original. Pour s'adapter aux nouvelles contraintes occasionnées par leur ancrage à la terre (immobilité), les plantes ont effectivement dû développer des réseaux métaboliques mettant en jeu des facteurs de transcription qui se distinguent remarquablement d'un point de vue de leur séquence des facteurs de transcription procaryotes et autres eucaryotes (Shiu et al., 2005).

A. thaliana contient de nombreuses duplications de gènes en tandem et duplications à large-échelle interchromosomiques qui représenteraient plus de 60% du génome. Ces duplications sont à l'origine de grandes familles de protéines qui peuvent se voir réparties à l'ensemble des cinq chromosomes. Parmi ces familles, deux ont été particulièrement amplifiées chez *A. thaliana* comparativement aux animaux et champignons, à savoir les protéines MYB (136 membres) et les protéines à boîte MADS (82 membres). Le DBD des protéines MYB repose sur une structure HTH pourvue de trois résidus tryptophane espacés régulièrement. Chez *A. thaliana*, la plupart de ces protéines MYB appartiennent à la classe R2R3-MYB (131 membres) contenant deux répétitions imparfaites du motif MYB, tandis que la classe R1R2R3-MYB, beaucoup plus classique chez les animaux, est très rare (5 membres; Riechmann et al., 2000). Le DBD des protéines à boîte MADS a déjà été détaillé quant à lui dans le paragraphe III.1.a.4 'les autres hélices α '.

Les facteurs de transcription spécifiques au règne végétal ont été classés et nommés en fonction de leur DBD. Une revue publiée en l'an 2000 fait état de ces DBD, décrits en onze larges familles: les protéines WRKY (72 membres), NAC (109 membres), AP2/ERF (144 membres), GARP, GRAS, Trihelix DNA, TCP, *Auxin response factors* (ARFs), SBP (17 membres), Nin-like et ABI3/VP1 (Riechmann et al., 2000; Yamasaki et al., 2008). De ces onze familles, beaucoup présentent des homologies structurales avec d'autres protéines eucaryotes: le DBD des WRKY et des NAC ressemble au domaine coordinateur de zinc GCM d'insectes, les protéines GRAS seraient apparentées aux protéines STATs, et des protéines Trihelix DNA et GARP auraient la même origine que les protéines MYB. Sur les quelques DBD présentés comme spécifiques aux plantes, peu 'survivent' à la comparaison avec des nouvelles entrées de structures animales et procaryotes dans la banque PDB. Les facteurs de transcription AP2 par exemple, l'une des trois familles de facteurs de transcription les plus abondantes chez *A. thaliana*, semblaient apporter un nouveau mode de reconnaissance de l'ADN par un feuillet β (Riechmann et al., 2000). Mais récemment, des homologies structurales ont été mises en évidence avec les intégrases du transposon procaryotique Tn916 et du bactériophage λ , ainsi qu'avec des endonucleases de levure et putativement de bactéries, de phages et de protistes (Yamasaki et al., 2008).

Ainsi, ces observations indiquent que la majorité des DBD spécifiques aux plantes dériveraient d'espèces non végétales, DBD qui ont ensuite été amenés à se dupliquer pour former des familles multigéniques chez les végétaux supérieurs.

2. Points sur les connaissances biochimiques relatives à LFY

Dans ce contexte, comment se positionne la protéine LFY ? Des recherches d'homologies de séquence n'ont permis de mettre en évidence aucune homologie apparente avec d'autres facteurs connus. LFY est donc un facteur de transcription spécifique aux plantes de 49kDa, soit 421 acides

aminés. Chez les Angiospermes, il est présent généralement en une copie, ou parfois deux, rarement plus. Si tel est le cas, ces copies supplémentaires ne semblent pas avoir acquis de subfonctionnalisation majeure (Frohlich, 2000; Himi et al., 2001). Chez le maïs, par exemple, il existe deux copies mais seul le double mutant possède un phénotype. Chez les Gymnospermes, deux paralogues, *LFY* et *NEEDLY (NLY)*, auraient un rôle partiellement divergent, nous y reviendrons dans le paragraphe IV.3. 'Théories sur l'apparition des fleurs'. Chez les Mousses et les Fougères, il existe plusieurs copies de *LFY* mais elles semblent étroitement apparentées et aucune évidence à ce jour ne suggère non plus de subfonctionnalisation majeure (Frohlich, 2000; Himi et al., 2001). Au bilan, contrairement aux autres facteurs évoqués dans le paragraphe précédent, *LFY* ne s'est pas dupliqué pour former une famille multigénique.

a) Régulation de l'expression d'*AP1*

Les premières données relatives à la protéine *LFY* ont été obtenues par Parcy et collaborateurs en 1998 à partir de l'expression chez *Escherichia coli* de protéines recombinantes délétées de leurs trente derniers acides aminés. Cette délétion d'après un essai d'expression constitutive dans les plantes n'entraîne qu'une modification minimale de l'activité de *LFY*. Ces auteurs ont purifié *LFY* en conditions dénaturantes par chromatographie d'affinité pour la double étiquette N et C-terminale de la protéine recombinante et ont montré qu'elle liait un oligonucléotide de 30pb correspondant à une séquence du promoteur d'*AP1*. Cet oligonucléotide possède une séquence pseudo-palindromique par rapport à une base médiane (Fig.12, soulignée). La mutation de deux bases de cette séquence (Fig.12, oligonucléotide *AP1 m1* en italique) entraîne la perte de liaison de la protéine à l'ADN, démontrant que *LFY* reconnaît spécifiquement l'oligonucléotide sauvage (Fig.13A). L'importance *in planta* de ce site de liaison de *LFY* au niveau du promoteur d'*AP1* n'a jamais été démontrée. Reyes Benlloch en stage post-doctoral dans notre équipe a travaillé sur cette question (données non publiées), j'y reviendrai dans la partie résultats.

b) Régulation de l'expression d'*AG*

En 1999, Busch et collaborateurs ont étudié l'activation d'*AG* par *LFY*. Ils ont pour cela introduit dans des plantes transgéniques le gène rapporteur *GUS* (codant pour la beta glucuronidase) sous le contrôle de versions tronquées des régions régulatrices d'*AG* (Busch et al., 1999). Ils ont ainsi délimité la région minimale de réponse à *LFY* au deuxième intron d'*AG*. *In vitro*, ils ont recherché plus précisément la séquence reconnue par *LFY* et montré par EMSA que *LFY* reconnaissait spécifiquement deux séquences, *AG-I* et *AG-II* (Fig.12 oligonucléotides *AG-I WT*, *m1*, *m2* et *AG-II WT*, *m1* et *m2*; Fig.13B), selon une affinité comparable dans les deux cas, mais qui serait inférieure à celle de *LFY* pour *AP1*. Ces deux sites *AG-I* et *AG-II* sont séparés de 47pb, et présentent comme le site *AP1* un motif pseudo-palindromique (Fig.12, en gras). L'importance biologique de ces deux sites a été validée

par l'étude *in planta* de l'effet de délétions au niveau de ces sites sur le profil d'expression d'AG (Busch et al., 1999). Ainsi, par alignement des séquences d'AG-I, AG-II et AP1, une séquence consensus CCANTG de liaison de LFY à l'ADN a pu être établie (Fig.12).

<i>AP1</i>	ACATTGTACGGACCACTGGTCCTTCCCCAA
<i>AG-I</i>	TAAATTTAATCCAATGGTTACAATTTTAAT
<i>AG-II</i>	GGATTTATACCCAATGTGTTAATGGGTTGT
Consensus	CCANTG
<i>AP3 site I</i>	CTTAAACCCTAGGGGTAATA
<i>AP3 site II</i>	TTCTATTTTCCAAGGATCTTTAGTTAAAGGC
Consensus	CCNNNG
<i>AP1 m1</i>	ACATTGTACGGACCACT TT TCCTTCCCCAA
<i>AG-I m1</i>	TAAATTTAATCCAAT AA TTACAATTTTAAT
<i>AG-I m2</i>	TAAATTTAATCCA CT GGTTACAATTTTAAT
<i>AG-II m1</i>	GGATTTATACCCAAT AA GTTAATGGGTTGT
<i>AG-II m2</i>	GGATTTATACCCA CT GTGTTAATGGGTTGT
<i>AP3 site Im</i>	CTTAAACCCTA TT ATTAATA
<i>AP3 site IIm</i>	TTCTATTTTCCA TT ATCTTTAGTTAAAGGC

Figure 12: Sites de liaison de LFY au niveau de ses gènes cibles AP1, AG et AP3

Seul le brin orienté 5'-3' est représenté ici. Les séquences de liaison palindromiques (nous nous permettons d'élargir la définition de ce terme ici où le palindrome n'est visible que si l'on considère les deux brins) sont indiqués en gras, la base médiane est soulignée, et les mutations sont représentées en rouge et en italiques.

Cependant, LFY n'est pas capable à lui seul d'activer AG et nécessite un co-régulateur. Lohmann et collaborateurs ont établi par EMSA en 2001 l'action concertée de LFY et WUS sur l'activation d'AG, à partir de protéines issues d'extraits bruts de levure *S. cerevisiae* (Fig.13C). Il est intéressant de noter que des sites reconnus par WUS sont systématiquement disposés à proximité des sites AG-I et AG-II reconnus par LFY bien que dans des sens opposés (Fig.14), et que l'action concertée de LFY et WUS entraîne une activation synergique d'AG. Ces données pourraient suggérer que LFY et WUS interagissent. Pourtant les résultats d'EMSA montrent que LFY et WUS contactent l'ADN indépendamment, et que le mélange des deux protéines ne crée pas de complexe supplémentaire (EMSA réalisé avec le site AG-II, Fig.12 et 13C). De plus, ces résultats ont été confirmés par expérience de coimmunoprécipitation où aucun hétérodimère LFY/WUS n'a pu être observé (Lohmann et al., 2001).

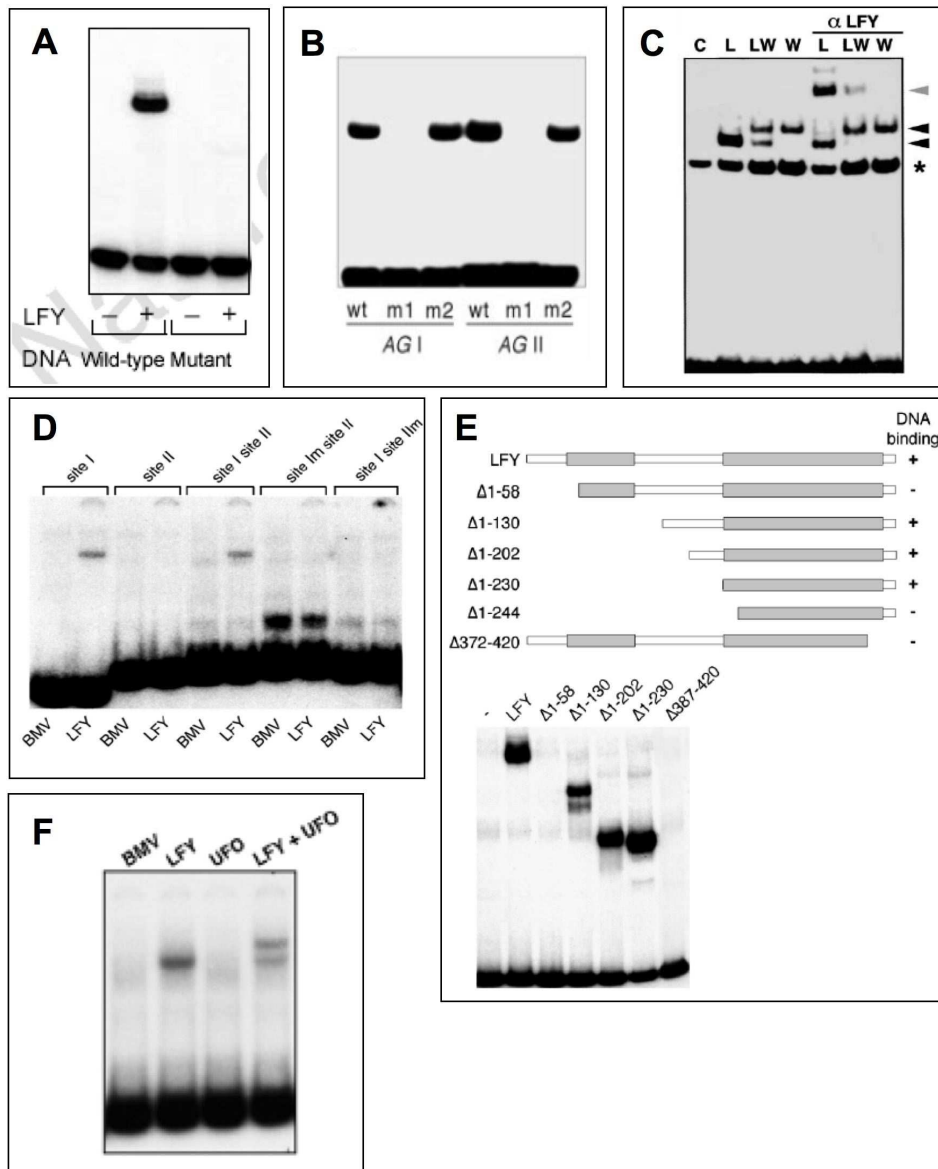


Figure 13: Récapitulatif des EMSA publiés à ce jour sur LFY

(A) EMSA réalisé avec AP1 WT et AP1 m1 (Parcy et al., 1998); (B) EMSA réalisé avec AG-I et AG-II WT ou mutés (Busch et al., 1999); (C) EMSA réalisé avec AG-II; C: contrôle, L:LFY, W:WUS, α LFY: anticorps anti-LFY; l'étoile signale un retard non spécifique; les triangles noirs les retards spécifiques de LFY et WUS, et le triangle gris le retard occasionné par le complexe LFY/anticorps anti-LFY (Lohmann et al., 2001); (D) EMSA réalisé avec les oligonucléotides AP3 site I, site II, site Im et site Ilm (précisé); BMV: Contrôle *Brome Mozaic Virus* exprimé par système TNT (Lamb et al., 2002); (E) EMSA réalisé avec AP1 WT et différentes truncations de LFY (précisé; Maizel et al., 2005); (F) EMSA réalisé avec AP3 site I (Chae et al., 2008).

En 2003, Hong et collaborateurs viennent réévaluer la liste des sites de liaison de LFY au niveau du deuxième intron d'AG suite à une étude phylogénétique entre orthologues de LFY chez 29 Brassicacées: LFY contacterait non pas deux, mais quatre sites (Hong et al., 2003). De la même manière que les sites I et II, AG-III se situe à proximité d'un site putatif de liaison de WUS (Fig.14). Le site AG-IV ne serait pas quant à lui à proximité d'un site de liaison de WUS. Selon le même principe de fusion rapportrice (AG::GUS) que Busch et collaborateurs (1999), ces auteurs proposent que le site AG-IV n'intervienne que dans l'expression précoce d'AG *in planta*. Les sites AG-I, AG-II, AG-III et les sites

de liaison de WUS correspondants seraient eux importants aussi bien à la mise en place qu'à la maintenance de l'expression d'AG au niveau des étamines et des carpelles.

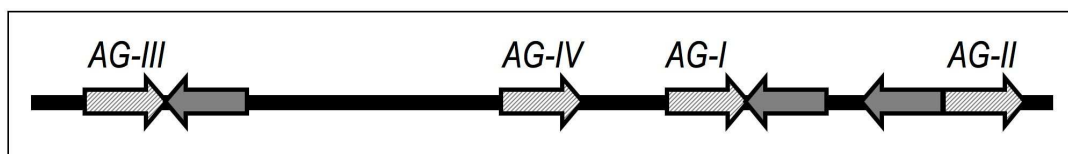


Figure 14: Représentation des sites de liaison de LFY et WUS au niveau du deuxième intron d'AG (d'après Busch et al., 1999 et Hong et al., 2003)

Sites de liaison de LFY représentés par des flèches rayées (site AG-I à -IV précisés), site de liaison de WUS représenté par des flèches grises.

c) Régulation de l'expression d'AP3

En 2002, Lamb et collaborateurs s'intéressent à l'activation d'AP3 par LFY. Selon le même principe que Busch et collaborateurs (1999), ils délimitent *in planta* un domaine minimal de réponse à LFY de 52pb dans le promoteur d'AP3. Après avoir coupé ce domaine en deux (sites I et II), ils ont constaté par EMSA à partir d'une protéine non purifiée produite en transcription traduction *in vitro* (TNT) que LFY liait spécifiquement le site I (Fig.13D). Étrangement, la mutation du site II, site qui par lui-même n'est pas reconnu par LFY, annule la liaison de LFY au site I. Les auteurs proposent pour expliquer ce constat que le site II intervienne dans la conformation de l'ADN. L'importance des sites I et II a ensuite été testée *in planta* selon le même principe que Busch et collaborateurs, où ici l'expression du gène rapporteur *GUS* est sous le contrôle de différentes versions du promoteur d'AP3 (constructions sauvage ou délétées du promoteur d'AP3). Étonnement, les délétions abolissant la liaison de LFY au niveau d'AP3 par EMSA n'ont pas d'effets clairs sur l'expression d'AP3 *in planta*. Pour l'expliquer, Lamb et collaborateurs proposent deux alternatives: soit ils n'ont finalement pas identifié le vrai site de liaison de LFY au niveau du promoteur d'AP3, soit l'induction d'AP3 *via* d'autres facteurs de transcription dont AP1 est suffisamment redondante pour pallier l'absence d'activation par LFY. Par alignement de séquences aux sites AG-I, AG-II et AP1, les sites AP3-I et AP3-II réduisent la séquence consensus de LFY à CCNNNG (Fig.12).

Ici encore LFY n'est pas capable à lui seul d'activer l'expression d'AP3 et agit pour cela avec UFO (Lee et al., 1997). En 2008, Chae et collaborateurs montrent par EMSA à partir de protéines produites en système TNT que LFY et UFO interagissent sur le promoteur d'AP3: UFO à elle seule ne peut lier l'ADN mais le mélange des deux protéines entraîne la formation d'un second complexe en plus de celui correspondant à LFY liée seul sur l'ADN (Fig.13F). Ce résultat est confirmé par expérience de double hybride chez la levure, où les auteurs suggèrent que l'interaction des deux protéines se fait *via* la région C-terminale de LFY. Cette interaction, observée également en expérience de coprécipitation GST ('GST pull down') ne dépend pas de l'ADN. Une de ses conséquences pourrait être l'ubiquitination de LFY par UFO (Chae et al., 2008).

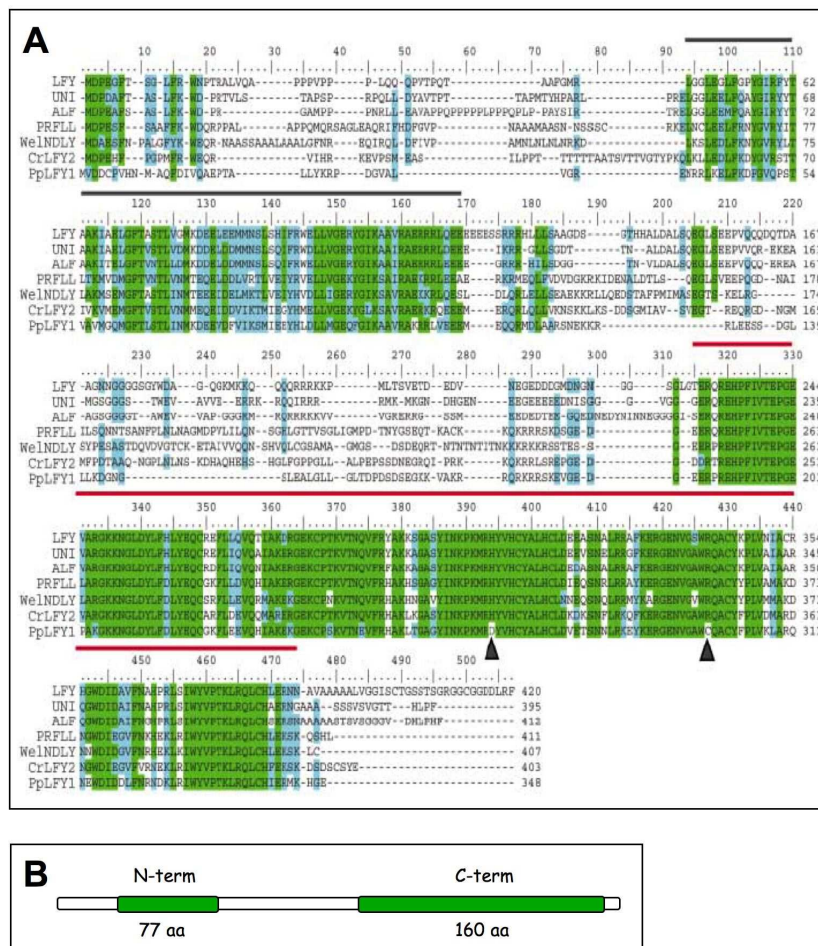


Figure 15: Régions conservées de la protéine LFY de différentes espèces

(A) Alignement de séquence de LFY (*Arabidopsis*) et 6 gènes orthologues chez le pois (*UNI*), le pétunia (*ALF*), le pin (*PRFLL*), *Welwitschia* (*WeINDLY*), la fougère *Ceratopteris* (*CrLFY2*) et le mouss *Physcomitrella patens* (*PpLFY1*) d'après Maizel et al., 2005. Les résidus colorés sont présents dans au moins 70% des séquences. En vert, résidus identiques; en cyan, résidus aux propriétés biochimiques identiques. Les flèches signalent les deux résidus variants chez la mouss *P.patens*; (B) représentation schématique des deux régions hautement conservées du gène LFY, et indication de leur nombre d'acides aminés (aa) chez *A.thaliana*. Le gène LFY entier contient 420 aa.

d) Evolution de LFY

En 2005, dans un article orienté vers l'étude 'évo/dévo' (synthèse de la biologie du développement et des théories de l'évolution), Maizel et collaborateurs mettent au jour à partir d'un alignement de séquences entre protéines homologues LFY d'espèces végétales très éloignées d'un point de vue évolutif (Mousses, Fougères, Gymnospermes et Angiospermes) deux régions extrêmement conservées, l'une de 77 aa en position N-terminale, l'autre de 160 aa en position C-terminale (Fig.15). Ces régions conservées sont potentiellement importantes pour la fonction des protéines LFY. En effet, dans le cas des protéines SBP par exemple, autres types de facteurs de transcription spécifique du règne végétal, seule une région est hautement conservée des Mousses aux Angiospermes, région correspondant au DBD (Riese et al., 2007). Effectivement, à partir de protéines d'*A. thaliana* tronquées produites en système TNT, cette équipe montre en EMSA que le domaine

minimal de liaison à l'ADN débute au résidu 230 (*i.e.* au ras de l'extrémité N-terminale de la région conservée C-terminale, Fig.15) et contient tout le domaine conservé C-terminal (Fig.13E).

De ces résultats sur l'étude biochimique de LFY, il ressort donc que LFY contacte l'ADN via son domaine conservé C-terminal, domaine qui interviendrait également dans l'interaction avec UFO. Aucune donnée n'est apportée quant au rôle que pourrait jouer le domaine conservé N-terminal. De plus, LFY reconnaîtrait une séquence consensus CCANTG^{G/T} en considérant les sites AP1 et AG-I, et AG-II, séquence qui se réduirait à CCNNNG en prenant également en considération les sites AP3-I et AP-3II.

Nous avons ainsi vu comment LFY régule les différentes étapes conduisant à l'élaboration de la fleur chez A. thaliana, rôle qui serait partiellement à totalement conservé chez les plantes à fleurs. Malgré ce rôle très spécifique, le gène LFY est déjà identifié chez les plantes dépourvues de fleurs (Gymnospermes, Fougères, Mousses). C'est à l'appui de ce constat que LFY a été intégré dans des théories de l'évolution visant à expliquer l'apparition des plantes à fleurs. Le chapitre qui suit fait état de cette réflexion.

IV. LFY, un acteur central de l'apparition des fleurs ?

1. « Un abominable mystère »

L'origine des fleurs est une véritable énigme qui, depuis plus d'un siècle, anime le monde de la biologie évolutionniste. Charles Darwin (1809-1882) qualifiait déjà cette apparition d'« abominable mystère ». Aujourd'hui, David Baum, chercheur-théoricien contemporain, continue à affirmer « *The origin of the angiosperms is to botany what the Cambrian explosion is to zoology* » (Baum and Hileman, 2006).

La fleur se caractérise par le rapprochement des organes reproducteurs mâles et femelles sur un même axe, la création du périanthe, le raccourcissement des entrenoeuds aboutissant à une structure compacte et la disposition des organes en couronne (Fig.16, Frohlich, 2000). Comprendre le mécanisme moléculaire à l'origine de ces innovations est un enjeu type de la stratégie 'évolution-développement' ('évo-dévo'). A ce compte, de nombreuses équipes s'attèlent à établir les liens de parenté existant entre les grands groupes végétaux afin de déterminer à quoi pouvait ressembler l'ancêtre des plantes à fleurs, question non sans difficulté puisque les Angiospermes n'ont pas d'antécédents évidents parmi les Gymnospermes vivants voire même fossiles.

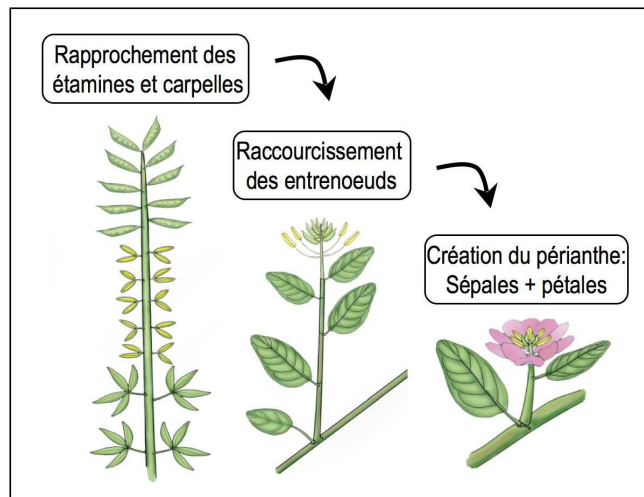


Figure 16: Transformations morphologiques à l'origine de la fleur (d'après Baum et Hileman, 2006)

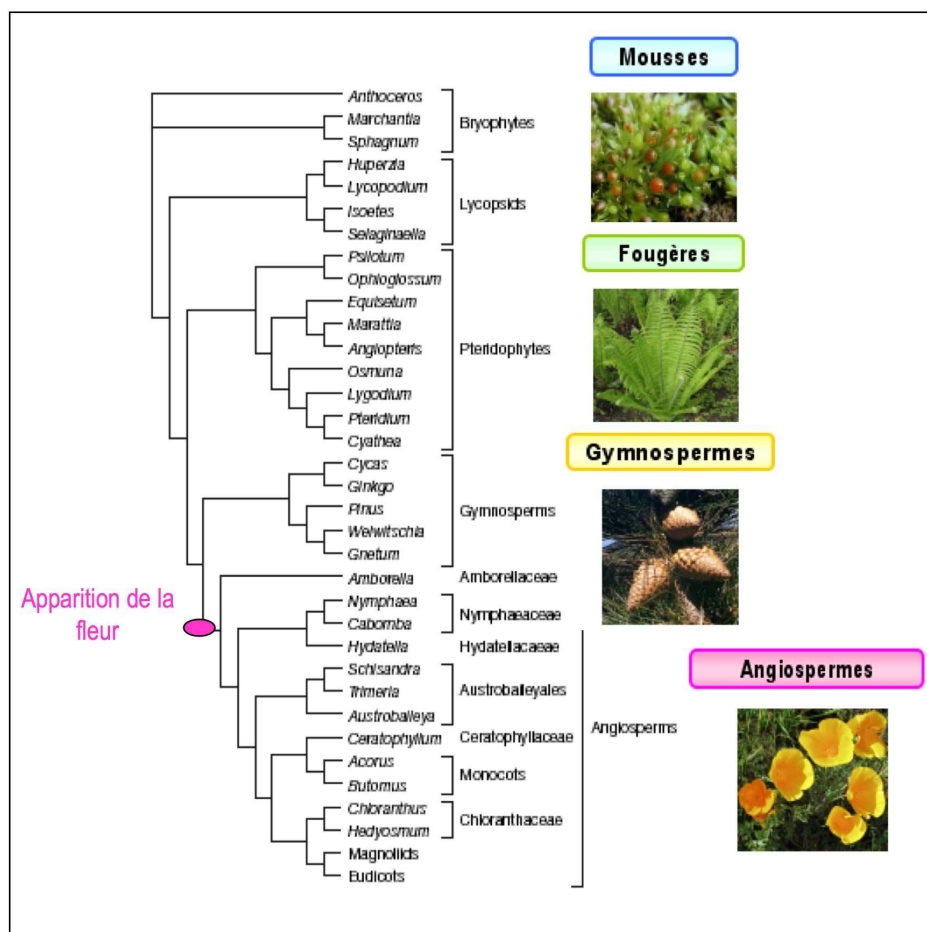


Figure 17: Arbre phylogénétique des groupes végétaux actuels (d'après Frohlich et Chase, 2007)

C'est Linné (1708-1778) qui fut le premier à proposer une classification des espèces végétales selon leurs propriétés florales et fruitières. Aujourd'hui, cet arbre phylogénétique s'est considérablement enrichi suite à un approfondissement des analyses morphologiques, notamment sur des fleurs fossiles, ainsi qu'à un apport de la phylogénie moléculaire. Cela alimente l'espoir d'une reconstruction

morphologique de l'Angiosperme primitif (Fig.17; Frohlich and Chase, 2007). D'ici là, les Angiospermes primitives ANA (*Amborellaceae-Nymphaeales-Austrobaileyales*; Fig.17) constituent un outil important. Chez ces espèces en effet, beaucoup de fleurs ont un nombre variable d'organes floraux qui sont souvent arrangés en spirale, suggérant que les ANA ont une homéostasie développementale moins élaborée que celle des Angiospermes supérieurs (Soltis et al., 2007). Etudier les mécanismes du développement floral chez ces espèces pourrait aider à bâtir un portrait moléculaire des premiers Angiospermes.

La logique de l'approche évo-dévo est de supposer que les réseaux de régulation contrôlant le développement des Angiospermes comme *A. thaliana* puissent avoir subi des modifications qui aient contribué à l'origine des plantes à fleurs. Le réseau impliquant LFY et ses gènes cibles ABC est un candidat de choix car il est à la base de la formation des méristèmes et des organes floraux. Une première approche consiste donc à identifier les homologues ou orthologues de ces gènes du réseau angiosperme et à étudier leurs profils d'expression chez les Gymnospermes.

2. Facteurs de transcription d'intérêt

Les protéines à boîtes MADS du modèle ABCE seraient des acteurs privilégiés à l'origine de l'apparition de la fleur. Nous avons vu en effet qu'elles spécifiaient les organes floraux et qu'elles étaient exprimées au niveau de zones correspondant typiquement à leurs zones d'action, point particulièrement intéressant car il facilite leur étude fonctionnelle. Que connaît-on ainsi de ces protéines chez les différents groupes végétaux ?

Les gènes 'B' et 'C' homologues des protéines angiospermes du modèle ABCE sont déjà présentes chez les Gymnospermes à la différence des gènes 'A' et 'E' qui eux seraient spécifiques aux Angiospermes (Frohlich and Chase, 2007; Zahn et al., 2005; Zahn et al., 2006).

Chez *A. thaliana*, les gènes 'A' s'expriment très tôt, dès le stade 1, et de façon uniforme au niveau des jeunes méristèmes floraux pour se limiter ensuite au primordia des sépales et pétales, en accord avec leur rôle temporel précoce dans l'identité des méristèmes floraux puis tardif dans l'identité des organes floraux. La fonction 'A' dans l'identité méristématique florale serait conservée chez tous les Angiospermes. En revanche, leur fonction tardive chez *A. thaliana* dans l'identité des sépales, et dans l'identité des pétales en collaboration avec les gènes 'B', semble propre aux espèces proches d'*A. thaliana* (Frohlich and Chase, 2007). Cette absence de mécanisme de détermination spécifique de l'identité des sépales chez des plantes éloignées d'*A. thaliana* s'accorde bien avec l'hypothèse d'un périanthe limité aux pétales chez la fleur originelle (Albert et al., 1998), et avec l'existence de tépales chez certaines Angiospermes, organes intermédiaires entre les sépales et les pétales qui viennent les remplacer au niveau de la fleur chez ces espèces.

La présence de deux gènes 'B' (*AP3/PI* chez *A. thaliana*, *DEF/GLO* chez *A. majus*) chez les Angiospermes résulterait d'un événement de duplication propre aux Angiospermes. Nous avons vu

qu'ils étaient exprimés au niveau des primordia des pétales et des étamines. Chez les Gymnospermes, l'unique gène 'B' est exprimé dans les structures mâles en cours de développement. Il semble donc que la fonction de ces gènes dans l'identité des organes reproducteurs mâles ait été conservée (Frohlich and Chase, 2007).

Les gènes 'C' d'Angiospermes s'expriment au niveau des primordia des étamines et des carpelles. Chez les Gymnospermes, les gènes 'C' sont exprimés aussi bien dans les structures mâles que femelles, suggérant là aussi une conservation fonctionnelle (Frohlich and Chase, 2007).

Ainsi, l'apparition des gènes 'A' et 'E' et ces conservations apparentes des fonctions des gènes 'B', 'C' des Gymnospermes aux Angiospermes sont à l'origine de théories sur l'apparition de la fleur.

En parallèle à l'étude de ces protéines à boîte MADS, un intérêt de plus en plus vif est entrain d'être porté à LFY. Malgré sa très grande conservation, les fonctions de LFY ont considérablement évolué au sein des grands groupes végétaux. Effectivement, il a été montré qu'initialement, chez la Mousse *Physcomitrella patens*, LFY était nécessaire à la mise en place de la première division zygotique (Tanahashi et al., 2005). Comment expliquer ainsi que LFY soit parvenu à acquérir un rôle de régulateur central de la floraison chez les Angiospermes ? Nous allons voir dans quelles mesures ses fonctions ont pu évoluer pour créer la fleur.

3. Théories sur l'apparition des fleurs

La grande difficulté de l'évo-dévo est de parvenir à émettre des théories simples, explicites, et testables. Je ne présenterai ici que les deux théories qui répondent au mieux à ces critères.

a) *Mostly Male*

La théorie '*Mostly Male*' part du constat que LFY d'Angiospermes possède deux paralogues chez les Gymnospermes, LFY et NEEDLY (NLY; Frohlich, 2000).

L'étude de l'expression des deux copies du pin (NLY et PRFLL; Mellerowicz et al., 1998; Mouradov et al., 1998) a laissé émerger l'idée que chez les Gymnospermes, LFY et NLY aient deux rôles distincts. PRFLL, orthologue de LFY, exprimé dans les territoires spécifiques des organes reproducteurs mâles, serait responsable de leur identité, alors qu'inversement, NLY, exprimé dans les territoires spécifiques des organes reproducteurs femelles, contrôlerait leur développement.

Les études ultérieures du profil d'expression de LFY et NLY, notamment chez les conifères *Picea*, *Podocarpus* et *Taxus*, n'ont pas confirmé cette observation d'expression spécifique du sexe des deux paralogues chez le pin (Vazquez-Lobo et al., 2007). Chez ces espèces, les orthologues de PRFLL et NLY sont exprimés aussi bien au niveau du développement précoce du cône femelle qu'au niveau du développement précoce des structures mâles. Néanmoins, ces profils d'expression ne s'opposent pas nécessairement à l'idée initiale, à savoir que LFY est impliqué dans l'identité des organes reproducteurs mâles, et NLY est impliqué dans l'identité des organes reproducteurs femelles. La distinction pourrait se

faire au niveau des gènes cibles de chacun. LFY pourrait réguler les gènes 'B', alors que NLY régulerait les gènes 'C'. D'après les connaissances sur le modèle ABC, le territoire mâle résulterait alors de l'activation des gènes 'B' et 'C' par les deux paralogues et le territoire femelle de l'activation des gènes 'C' par NLY.

La théorie *Mostly Male* prédit que la structure bisexuée s'est formée à partir du cône mâle d'où elle tire son nom ('principalement mâle'). LFY serait donc devenu capable de réguler à la fois les gènes 'B' et 'C' en des territoires parfois chevauchant (identité mâle) parfois distinct (identité femelle) convertissant le cône mâle en en structure florale primaire. NLY, devenu inutile aurait été alors perdu.

Cette théorie est étayée par d'autres observations: le fait que le mutant nul *lfy* ne soit plus capable de produire d'étamines alors qu'il parvient toujours à produire des carpelles supporte l'idée que LFY est absolument requis au bon développement des organes reproducteurs mâles mais que des mécanismes indépendants de LFY sont redondants dans l'identité femelle (Weigel et al., 1992). De plus, la formation ectopique d'ovules au niveau de feuilles de *Ginkgo biloba* notamment suggère que les ovules ectopiques sont relativement faciles à produire, en accord avec l'idée qu'il est plus probable que la fleur dérive d'une structure mâle originelle à laquelle se serait ajoutée la structure reproductrice femelle qu'inversement (Frohlich, 2000). Enfin, cette théorie est cohérente avec la conservation supposée entre Gymnospermes et Angiospermes de l'activité 'B' dans l'identité mâle.

b) Out of Male/Out of Female

La théorie '*Out of Male/Out of Female*' se focalise sur le caractère bisexué des fleurs. Theissen et collaborateurs ont observé que des conifères modernes produisaient parfois des cônes bisexués. Une perte d'expression des gènes 'B' à la pointe des cônes mâles gymnospermes pourrait occasionner une invasion de l'identité femelle à ce niveau, ou inversement, l'expression ectopique des gènes 'B' à la base des cônes femelles pourrait induire la présence de l'identité mâle à ce niveau. Le caractère bisexué des fleurs pourrait alors aussi bien venir de structures ancestrales mâles que femelles (Theissen and Melzer, 2007).

Baum et Hileman (2006) apportent d'autres éléments sur le mécanisme de la théorie '*Out of Male*' (Baum and Hileman, 2006). Ils suggèrent qu'une très forte expression des protéines responsables de l'activité 'C' dans la partie centrale des cônes mâles peut être perturbée par les gènes 'E'. Les produits de l'activité 'E' interagiraient avec les produits de l'activité 'C', et neutraliseraient par conséquent l'interaction entre les produits de l'activité 'B' et 'C': l'identité femelle l'emporterait.

c) Premières caractérisations évolutives de la protéine LFY

Pour valider ces théories, une étude méticuleuse du réseau de régulation reliant LFY aux gènes MADS sera donc nécessaire. Même si l'analyse de l'expression de ces gènes chez des espèces très

éloignées d'un point de vue évolutif est un pré-requis indispensable, étudier l'évolution de la fonction des protéines elles-même sera l'outil incontournable à la compréhension de ces mécanismes.

Dans cette optique, Maizel et collaborateurs (2005) ont proposé une étude évolutive de LFY. Ils ont pour cela effectué une série de complémentations fonctionnelles d'un mutant *lfy* d'*A. thaliana* avec des homologues de différentes espèces. La complémentation se révèle d'autant moins bonne que les espèces sont éloignées d'*A. thaliana* d'un point de vue évolutif. En effet, si les gènes provenant des Gymnospermes *PRFLL* (pin) et *WelNDLY* (*Welwitschia mirabilis*) parviennent à restaurer partiellement le phénotype sauvage des fleurs d'*A. thaliana*, les gènes provenant des Fougères *CrLFY2* (*Ceratopteris richardii*) et *AILFY4* (*Angiopteris lygodiifolia*) sont encore moins efficaces. Finalement, les homologues les plus basaux, ceux de la Mousse *P. patens* (*PpLFY1* et *PpLFY2*), sont totalement inactifs chez *A. thaliana*.

Ces auteurs ont ensuite étudié par microarray la capacité de ces différents orthologues de LFY à induire l'expression de gènes cibles chez *A. thaliana*. Là encore, sur les 16 gènes les plus induits suite à l'activation du gène de référence LFY, plus on s'éloigne d'un point de vue évolutif des Angiospermes supérieurs, plus le nombre de gènes induits par les gènes homologues de LFY est faible, jusqu'à être nul chez la Mousse. Il est intéressant de noter que dans ce dernier cas, l'analyse plus exhaustive des résultats de microarray, sans se limiter à ces 16 gènes, ne révèle aucune activation de gènes suite à l'activation du gène LFY de Mousse, suggérant que la protéine qu'il code pourrait ne pas être capable de lier l'ADN.

Pour essayer d'identifier les modifications évolutives de la protéine LFY à l'origine de ces modifications fonctionnelles, les auteurs ont eu recours à l'étude fonctionnelle de protéines chimériques entre LFY et des orthologues d'autres espèces.

Dans un premier temps, ils ont réalisé différentes fusions traductionnelles entre régions conservées et non conservées de LFY et de CrLFY2 de Fougère. Leurs résultats révèlent que les modifications évolutives sont essentiellement contrôlées par les régions conservées, et que la région C-terminale impliquée dans la liaison à l'ADN jouerait un rôle prépondérant dans la divergence de fonction entre LFY et CrLFY2.

Ils ont ensuite observé que deux résidus hautement conservés varient dans la région conservée C-terminale entre LFY et son homologue *PpLFY1* chez la Mousse *P. patens* (chez *PpLFY1*, H312D et C345R, correspondant respectivement à H394D et C427R de la numérotation de Maizel et al.; fig.15A). Ils ont donc voulu savoir si une mutation de chacun de ces deux résidus chez *PpLFY1* en leur équivalent chez LFY pouvait modifier leur activité. Si la mutation R345C ne modifie pas les propriétés fonctionnelles de *PpLFY1*, la mutation D312H lui permet d'activer partiellement la transcription d'un gène cible connu de LFY (*AP1*) et de compléter partiellement le phénotype *lfy*

chez *A. thaliana*. Leurs résultats suggèrent donc que ce résidu His312 puisse jouer un rôle crucial dans la fonctionnalité de LFY.

Ces données préliminaires constituent une bonne base de travail en apportant une cartographie grossière des régions clés impliquées dans l'évolution fonctionnelle de LFY. Cependant, en l'absence de données structurales, il est difficile d'affiner ces résultats par l'étude 'à l'aveugle' des déterminants clés, les acides aminés. En effet, même si plusieurs acides aminés divergents d'un groupe végétal à l'autre sont révélés par alignement de séquence, leur analyse systématique sans idée de leur rôle potentiel deviendrait très rapidement laborieuse.

V. Objectifs de mon travail de thèse

LFY est un régulateur central de la floraison chez différentes Angiospermes. Il intervient dans les deux étapes clés de ce processus, à savoir l'intégration des stimuli exogènes et endogènes contrôlant le moment de la floraison et la détermination de l'identité du méristème floral et des organes floraux. En plus de ces fonctions décrites chez *A. thaliana*, plusieurs théories de l'évolution lui attribuent un rôle à l'origine de l'apparition des fleurs. Pourtant, la caractérisation biochimique et fonctionnelle de ce facteur original reste pauvre.

Le projet de ma thèse est donc d'étudier par approche biochimique et structurale la protéine LFY afin de mieux comprendre son mécanisme moléculaire d'action.

Les questions qui nous intéressent sont les suivantes:

- A quelle classe de facteur LFY appartient-il ? Ressemble-t-il au niveau structural à des facteurs déjà décrits ou est-il complètement original ? Est ce que sa structure peut nous aider à comprendre son origine et son fort niveau de conservation ?
- Comment reconnaît-il spécifiquement certaines séquences d'ADN ? Quels sont les acides aminés impliqués et quel est son mode de liaison (monomérique, multimérique) ?
- Est ce que son mode d'action et d'interaction avec ses partenaires peut nous aider à comprendre les bases moléculaires de la transition florale ?
- Comment LFY a-t-il évolué ? Est ce que des acides aminés au rôle crucial ont changé au moment de l'apparition de la fleur ? Doit-on s'attendre à des changements de spécificité de liaison ?

MATERIEL

1. Matériel bactérien

Les clonages classiques ont été effectués dans des bactéries *Escherichia Coli* de souche DH5 α F ϕ 80 *lacZ* Δ M15 Δ (*lacZYA-argF*)U169 *deoR* *recA1* *endA1* *hsdR17*(r_K^- , m_K^+) *phoA* *supE44* *thi-1* *gyrA96* *relA1* λ^- (Invitrogen) ou TOP10 F $^-$ *mcrA* *D*(*mrr-hsdRMS-mcrBC*) *f80lacZDM15* *DlacX74* *deoR* *recA1* *araD139* *D*(*ara-leu*)7697 *galU* *galK* *rpsL* (*StrR*) *endA1* *nupG* (Invitrogen). Les clonages Gateway (avant recombinaison) ont été effectués dans des bactéries *E. coli* souche DB3-1 F $^-$ *gyrA462* *endA1* *glnV44* Δ (*sr1-recA*) *mcrB* *mrr* *hsdS20*(r_B^- , m_B^-) *ara14* *galk2* *lacY1* *proA2* *rpsL20*(*Smr*) *xyl5* Δ *leu* *mtl1* (Invitrogen).

Les expressions de protéines ont été effectuées à partir de bactéries *E. coli* souche RosettaBlue(DE3)pLysS F $^-$ *ompT* *hsdS_B*(r_B^- m_B^-) *gal* *dcm* λ (DE3 [*lacI* *lacUV5-T7 gene 1* *ind1* *sam7* *nin5*]) *pLysSRARE* (*Cam^R*; Novagen) ou *E. coli* souche Rosetta2(DE3)pLysS F $^-$ *ompT* *hsdS_B*(r_B^- m_B^-) *gal* *dcm* (DE3) *pLysSRARE2* (*Cm^R*; Novagen). Des essais ont également été entrepris en *E. coli* BL21(DE3)pLysS F $^-$ *ompT* *gal* *dcm* *lon* *hsdS_B*(r_B^- m_B^-) λ (DE3) *pLysS*(*cm^R*; Novagen).

Les essais d'expression transitoire en tabac *Nicotiana benthamiana* sont réalisés à partir de transformations dans la souche C58C1 pMP90 d'*Agrobacterium tumefaciens* (Koncz and Schell, 1986).

2. Levure

Les expériences d'induction chez la levures ont été réalisées à partir de la souche *Saccharomyces cerevisiae* EGY48 *Mat α* *his3* *leu2::3LexAop-LEU2* *ura3* *trp1* *LYS2* (Invitrogen).

3. Matériel végétal

Les plantes *A. thaliana* F0 hémizygotes *lfy-5-VP16/.*, homozygotes *lfy-12/lfy-12* accession *Columbia* nous ont gentiment été données par D. Weigel (Institut Max Planck, Tuebingen, Allemagne). Les plantes *lfy-5-VP16/lfy-5-VP16*, *lfy-12/lfy-12* ont été obtenues par autofécondation des plantes F0, tandis que les plantes *lfy-5-VP16/lfy-5-VP16*; *LFY/LFY* ont été obtenues en F2 par croisement des plantes F0 à des plantes sauvages d'accession *Columbia* puis autofécondation des plantes F1. Les plantes homozygotes pour le transgène *lfy-5-VP16/lfy-5-VP16* ont été confirmées sur le critère d'une descendance 100% résistantes à la kanamycine (25 mg/L) en milieu 1/2MS. Les plantes homozygotes *lfy-12/lfy-12* ont été identifiées par observation du phénotype *lfy* (Huala and Sussex, 1992) et confirmées au même titre que les plantes homozygotes *LFY/LFY* par génotypage (cf Méthodes-Génotypage des plantes *lfy-12*).

L'allèle mutant *lfy-28* d'*A. thaliana* accession *Landsberg erecta* nous a également été donné par D. Weigel (Institut Max Planck, Tuebingen, Allemagne) et a été originellement isolé par J.C. Fletcher (PGEC, Albanie).

METHODES

1. Solutions d'usage courant

LB (Luria Bertani) préparation manuelle: 1% [m/v] bacto tryptone, 0.5% [m/v] extrait de levure, 1% [m/v] NaCl, pH7.0, + optionnel 1% agar pour obtenir un milieu solide

LB commercial: 2% [m/v] poudre LB Broth Base Lennox (Invitrogen) + optionnel 1% agar pour obtenir un milieu solide

LB liquide 0,5M NaCl: 2% [m/v] poudre LB Broth Base Lennox (Invitrogen), 2.42% [m/v] NaCl

Laemmli 1X: 2.5% β -mercaptoethanol, 2% SDS, 80mM Tris pH6.8, 10% glycérol, 0.004% bleu de bromophénol

PBS (Phosphate-Buffered Saline) 1X: 0.8% [m/v] NaCl, 0.02% [m/v] KCl, 0.144% [m/v] Na₂HPO₄, 0.024% [m/v] KH₂PO₄, pH7.4

TAE (Tris Acetate EDTA) 1X: 40mM Tris, 20mM acide acétique, 1mM EDTA pH8.0

TAE modifié 1X: 40mM Tris, 20mM acide acétique, 0.1mM EDTA pH8.0

TB (Terrific Broth) liquide préparation maison (protocole initial): 1.2% [m/v] Bacto Tryptone, 2.4% [m/v] extrait de levure, 1% [m/v] NaCl, 0.04% [v/v] Glycérol, 50 mM KPO₄ pH7.2

TB liquide commercial: 4.76% [m/v] poudre TB modified (Sigma), 0.8% [v/v] glycérol

TBE 10X: 10.8% [m/v] Tris, 5.5% [m/v] acide borique, 20mM EDTA 0.5M pH8.0

TBS (Tris Buffered Saline) 1X: Tris 50mM pH7.5, NaCl 100mM

TE (Tris EDTA): Tris 10mM pH8.0, EDTA 1mM

2. Biologie moléculaire

a) BiFC

Quatre combinaisons 35S::LFY-BiFC ont été réalisées en système Gateway par F. Parcy. Les constructions 35S::GFP (protéine fluorescente verte), *Act11::H2B-mRFP1* (marqueur nucléaire), et 35S::p19 (suppresseur de silencing) nous ont été gracieusement données respectivement par D. Orsaez (Université Polytechnique de Valencia, Espagne), F. Berger (RDP, Lyon, maintenant à Singapour) et D. Baulcombe (Centre John Innes, Norwich, Royaume-Uni). La construction 35S::GFP-LFY (XW19(N)) a été décrite par Wu et al., 2003.

b) Vecteurs d'expression en bactéries

Remarque: Les oligonucléotides mentionnés sont reportés au sein du tableau A. Toutes les enzymes de restriction proviennent du fournisseur New England BioLabs. Les vecteurs ont été vérifiés par séquençage.

Tableau A: Récapitulatif des oligonucléotides utilisés en biologie moléculaire

Oligonucléotide	Séquence
oCH1003	5'-GATCCGCTAGGCCTCGTCGC
oCH1004	5'-TCGAGCGACGAGGCTAGCG
oCH1007	5'-CATCTCGAGGTTCAAGAGCAAGCG
oCH1008	5'-GCGCCATGGGCATGATCCTGAAGTTTCACGA
oCH1009	5'-AGCCTCGAGGAAACGCAAGTCGTCGCCG
oCH1022	5'-CTCCATGGCGTTTATCGTAACGGAGCCTG
oCH1023	5'-CTCTCGAGCTAGTTCCGCTCCAAATGGCAAAG
oCH1024	5'-CTCTCGAGCTACAGCTTTGTTGAACATAC
oCH1025	5'-AGCGAGTTACATAAACGCGCCTAAAATGCGACACT
oCH1026	5'-CCACCAAGGTGACGGCCCAAGTATTCAGGTACGC
oCH1027	5'-GGACAGAGAGACAGGCGGAGCATCCGTTTATCGT
oCH1028	5'-GAGTTACATAAACAAGGCTAAAATGCGACACTACG
oCH1029	5'-CAGACAATTGCTAAAAAGCGTGGCGAAAAATGC
oCH1030	5'-AACCGCATTGTTGCGCTCCAAAGCGCAAAGCTGACG
oCH1031	5'-GTTCCGCTCCAAAGCGCAAAGCTGACGCAG
oCH1032	5'-GAGTTACATAAACAAGCTTAAAATGCGACACTACG
oCH1033	5'-ATCCATGGATCGCCAGACCAGCAGCAGGT
oCH1034	5'-CTTCTTCTGACGCTCACGAGCCTT
oFP1212	5'-GATCAGGTGGGGATCCCTCGAGTGGCCATCACCATCACCATCACTGA
oFP1213	5'-TCGATCAGTGATGGTATGGTATGGCCACTCGAGGGATCCCCACCT
oFP1215	5'-NNCTCGCACTGTTACGCTCTCCACTGCCTA
oFP1227bis	5'-GACAGAGGGAGCATCTGTTTATCGTAACGGAGCCCGGG
oFP1228bis	5'-CCCGGGCTCCGTTACGATAAACAGATGCTCCCTCTGTC
oFP1231	5'-CACCATGGAGCCGCCACAGCATCAGC
oFP1241	5'-CTCCATGGGTTTGGGACAGAGACAG
oFP1242	5'-CTCTCGAGCCCGGGCTAGAAACGCAAGTCGTCGCC
oFP1243	5'-CTCTCGAGCCCGGGCTATCCGGTACAGCTAATACCGCC
oFP1244	5'-CTCCATGGATAACGGCAACGGAGGTA
oFP1248	5'-GCAACCGCATTGTTCTTCTCCAAATGGCAAAGC
pETH1005	5'-GCAACGGAGGTAGTGGTTTGGGACAGAG

Milieu de sélection

Les bactéries poussent sur milieu liquide ou solide LB ou TB à 37°C auquel sont ajoutés des antibiotiques en fonction de la nature de la souche et du vecteur qu'elle reçoit. Les bactéries transformées avec un vecteur

dérivant des vecteurs pGEX-4T1 et pIVEX2.4d sont sélectionnées sur carbenicilline 100µg/mL, les vecteurs dérivés des vecteurs pET21b, pET30a, pETM-11, pETM-30a+, pETtrx1a et pETtrx1b sont sélectionnés sur kanamycine 50µg/mL. Pour les souches bactériennes *E. coli* RosettaBlue(DE3)pLysS, Rosetta2(DE3)pLysS et BL21(DE3)pLysS, est ajouté du chloramphenicol 34µg/mL.

Domaine C-terminal de LFY

Tous ces vecteurs sont dérivés du plasmide pETM-11 qui permet l'expression d'une protéine de fusion étiquetée par 6 histidines (Dummler et al., 2005). Le clonage de la construction LFY-C sauvage a été décrit dans Hamès et al., se référer à l'annexe 1. Les constructions LFY-C mutantes ont été obtenues par stratégie mégaprimé (Kirsch and Joly, 1998) à partir du vecteur pCH28 d'expression de 6his-LFY-C et du vecteur pETH19 (vecteur dérivé de pCH28 délété de son site de restriction *HindIII*, travail d'E. Thevenon). Brièvement, cette stratégie consiste à amplifier un fragment en introduisant une mutation et ensuite à utiliser le fragment comme primer ce qui incorpore la mutation dans le vecteur d'expression. L'ADN polymérase fidèle Phusion (Finnzymes) est utilisée pour la première étape et l'ADN polymérase fidèle Ex Taq (Takara) pour la deuxième. Chaque construction est générée à l'aide d'un couple d'oligonucléotides approprié: oCH1025/oFP1242 pour pCH45 (K307A), oCH1026/oFP1242 pour pCH46 (N291A), oCH1027/oFP1242 pour pCH47 (R237A), oCH1028/oFP1242 pour pCH48 (P308A), oCH1029/oFP1242 pour pCH49 (D280K), oCH1030/oFP1244 pour pCH50 (H387A R390A), oCH1031/oFP1244 pour pCH54 (H387A), oCH1032/oFP1244 pour pCH55 (*lfy-28*, P308L) et oETH1005/oFP1248 pour pEDW127 (R390A; travail d'E. Moyroud).

Le vecteur pCH56 (*lfy-5*, P240L) a été obtenu par stratégie de mutagenèse 'Quick Change' (Stratagene) selon les recommandations du fournisseur.

Les vecteurs pETH21 et pETH23 (respectivement *lfy-4* (E238K) et *lfy-20* (N306D); travail d'E. Thévenon) ont été obtenus par stratégie mégaprimé (Kirsch and Joly, 1998) à partir du couple d'oligonucléotides oFP1242/oFP1244 et des vecteurs pMX182 et pMX184.

Délimitation du domaine minimal de liaison à l'ADN:

Les vecteurs **pCH26**, **pCH27**, **pCH29**, **pCH42**, **pCH43** et **pCH44** (Fig.18) ont été obtenus selon le même protocole que pCH28 à partir respectivement des couples d'oligonucléotides oFP1241/oFP1242, oFP1241/oFP1243, oFP1243/oFP1244, oCH1022/oFP1242, oCH1023/oFP1244 et oCH1024/oFP1244.

Mutagenèse des cystéines et histidines conservées:

Les vecteurs **pCH33**, **pCH34**, **pCH35**, **pCH36**, **pCH37** et **pCH38** (mutation des histidines) et les vecteurs **pEDW10**, **pEDW11**, **pEDW12**, **pEDW13** et **pEDW14** concernant (mutation des cystéines) ont été obtenus par stratégie de mutagenèse 'Quick Change' (Stratagene) à partir de la matrice pCH28 (travail en collaboration avec E. Moyroud).

Autres vecteurs d'expression de LFY-Cterm sauvage testés:

LFY-Cterm-6his/vecteur d'origine pIVEX2.4d: les produits de digestion *NcoI/XmaI* issus des vecteurs pCH26, pCH27, pCH28 et pCH29 ont été introduits en *NcoI/XmaI* du vecteur pIVEX2.4d (Roche). Les vecteurs obtenus sont respectivement **pCH22**, **pCH23**, **pCH24** et **pCH25**.

LFY pleine longueur (LFY FL, Full Length)

- 1) GST-thrombine-LFY-6his/vecteur d'origine pGEX-4T1 modifié: pGEX-4T1 (Pharmacia Biotech) comporte une étiquette Glutathione S-transférase (GST) en 5'. Une étiquette 6 histidines (6his) a été introduite en 3' au niveau des sites *BamHI/XhoI* de ce vecteur par insertion de deux oligonucléotides complémentaires oFP1212 et oFP1213 pour obtenir pCH1. L'ADNc LFY sans codon stop a été introduit en deux fragments *BamHI/HindIII* et *HindIII/XhoI* de taille respective 970 pb et 330 pb, en 3' de l'étiquette GST et 5' des 6his de pCH1, par digestion *BamHI/XhoI*, permettant d'obtenir **pCH2**.
- 2) GST-thrombine-GFP-LFY-6his/vecteur d'origine pGEX-4T1 modifié: l'ADNc du gène codant pour la GFP a été introduit dans le vecteur pCH1 pour obtenir pFP138 (travail de F.Parcy). Le vecteur pFP138 a été converti en vecteur Gateway pCH7 (GST-Thrombine-GFP-6his) par insertion de la cassette RfB par digestion *StuI*. **pCH8** a été obtenu par réaction LR à partir du vecteur pENTRY-LFY et du vecteur pCH7.
- 3) GST-thrombine-LFY-GFP-6his/vecteur d'origine pGEX-4T1 modifié: l'ADNc codant la fusion traductionnelle GFP:LFY (*GFP* insérée dans l'ADNc de LFY) a été dessiné par X. Wu (vecteur pBS-GLFY; Wu et al., 2003). Cette fusion a été introduite dans le vecteur pCH2 par digestion *BamHI/HindIII* (travail de F.Parcy) pour obtenir **pJUL1**.
- 4) 6his-LFY-6his/vecteur d'origine pET21b modifié: le vecteur pET21b (Novagen) a été modifié par ajout d'une étiquette 6his en 5' et converti en vecteur Gateway par insertion de la cassette RfA pour obtenir le vecteur

pHisGTWHis (travail de J. Gaillard). **pCH15** a été obtenu par réaction LR à partir du vecteur pENTRY-LFY et du vecteur pHisGTWHis.

5) 6his-LFYtronquée-6his/vecteur d'origine pET21b modifié: **pCH17** a été obtenu par réaction LR à partir du vecteur pENTRY-LFYtronquée et du vecteur pHisGTWHis. Le vecteur pENTRY-LFYtronquée est dérivé de pENTRY-LFY par délétion de la région hydrophobe C-terminale (32 derniers acides aminés). Cette délétion améliorerait l'expression de LFY (Parcy et al., 1998).

6) 6his-LFY-6his/vecteur d'origine pETM-11: l'ADNc LFY a été amplifié par PCR avec les oligonucléotides oCH1008 et oCH1009 puis sous-cloné dans pGEM-T easy (Promega) pour obtenir pCH32bis. L'insert obtenu par digestion NcoI/XhoI de pCH32bis a été introduit en NcoI/XhoI de pETM-11 (Dummler et al., 2005) pour obtenir **pCH32**.

7) 6his-GST-TEV-LFY-6his/vecteur d'origine pETM-30: le vecteur **pETH18** a été obtenu par insertion du produit de digestion NcoI/XhoI de pCH32 en NcoI/XhoI de pETM-30 (collection EMBL, travail d'E. Thevenon).

8) 6his-GST-TEV-LFY/vecteur d'origine pETM-30: le vecteur **pCH52** a été obtenu par stratégie mégaprimer (Kirsch and Joly, 1998) à partir des oligonucléotides oFP1215 et oFP1243, de la matrice pETH18 pour la première PCR et pETM-30 pour la seconde.

9) 6his-trx carrier-TEV-LFY/vecteur d'origine pETtrx1a: le vecteur **pCH56** a été obtenu par insertion du produit de digestion NcoI/XhoI de pCH52 en NcoI/XhoI de pETtrx1a.

10) trx carrier-6his-TEV-LFY/vecteur d'origine pETtrx1b: le vecteur **pCH57** a été obtenu par insertion du produit de digestion NcoI/XhoI de pCH52 en NcoI/XhoI de pETtrx1b.

Homéodomaine de WUS (WUS-Hd)

1) 6his-trx carrier-TEV-WUS-Hd/vecteur d'origine pETtrx1a: le domaine WUS-Hd a été amplifié par PCR à partir des oligonucléotides oCH1033 oCH1034 et un vecteur contenant l'ADNc WUS (pCH53, cf ci-dessous) a été sous-cloné en vecteur pCR Blunt (Invitrogen) pour obtenir pCH60. Le vecteur **pCH61** est obtenu par insertion du produit de digestion NcoI/XhoI de pCH60 en NcoI/XhoI de pETtrx1a.

2) trx carrier-6his-TEV-WUS-Hd/vecteur d'origine pETtrx1b: le vecteur **pCH62** est obtenu par insertion du produit de digestion NcoI/XhoI de pCH60 en NcoI/XhoI de pETtrx1b.

WUS pleine longueur (WUS FL)

1) 6his-WUS-6his/vecteur d'origine pET30a+: le vecteur pET30a+ (Novagen) a été converti en vecteur Gateway par insertion de la cassette RfA (travail de J. Gaillard). Le vecteur **pCH6** a été obtenu par réaction LR entre ce vecteur pET30a+RfA et pENTRY-WUS.

2) GST-thrombine-WUS-6his/vecteur d'origine pGEX 4T1 modifié: un site StuI a été introduit en BamHI/XhoI de pCH1 par insertion de deux oligonucléotides complémentaires oCH1003 et oCH1004 pour obtenir pCH3. La cassette RfB a été clonée au niveau de StuI pour convertir pCH1 (GST-ADNc-6his) en vecteur Gateway, nommé pCH4. Le vecteur **pCH10** a été obtenu par réaction LR à partir de pENTRY-WUS et pCH4.

3) 6his-WUS-6his/vecteur d'origine pETM-11: un site XhoI inclus dans l'oligonucléotide oCH1007 a été introduit à l'extrémité 5' de WUS par PCR avec les oligonucléotides oCH1007 et oFP1231 sur la matrice pENTRY-WUS. Le produit PCR a été ensuite sous-cloné en pCR Blunt (Invitrogen) pour obtenir le vecteur pCH31bis. Le vecteur **pCH31** a été obtenu par insertion du produit de digestion NcoI/XhoI de pCH31bis en NcoI/XhoI de pETM-11.

4) 6his-GST-TEV-WUS/vecteur d'origine pETM-30: le vecteur **pCH53** a été obtenu par insertion du produit de digestion NcoI/XhoI de pCH6 en NcoI/XhoI de pETM-30.

5) 6his-trx carrier-TEV-WUS/vecteur d'origine pETtrx1a: le vecteur **pCH58** a été obtenu par insertion du produit de digestion NcoI/XhoI de pCH53 en NcoI/XhoI de pETtrx1a.

6) trx carrier-6his-TEV-WUS/vecteur d'origine pETtrx1b: le vecteur **pCH59** a été obtenu par insertion du produit de digestion NcoI/XhoI de pCH53 en NcoI/XhoI de pETtrx1b.

Protéase TEV (*tobacco etch virus*)

Les vecteurs d'expression His-TEV et His-TEV *point mutation* (mutation stabilisant la protéine) nous ont été gracieusement donnés par Ethan Lee (Université de Vanderbilt, Nashville, USA; communications personnelles) et le vecteur d'expression TEV His dérivant de pET29 nous provient d'Anne Houdusse (Institut Curie, Paris).

c) Vecteurs d'expression chez la levure

Pour l'expérience de coexpression *Ify:VP16/LFY*, les vecteurs d'expression pMX181 (*Ify-3-VP16*), pMX182 (*Ify-4-VP16*), pMX183 (*Ify-9-VP16*), pMX184 (*Ify-20-VP16*) produits par M. Busch (MPI Tuebingen) et pRHO27/1 (*Ify-5-VP16*; MPI Tuebingen) dérivent du vecteur initial p424 (expression inductible au galactose;

prototrophie pour le tryptophane; Mumberg et al., 1995). Le vecteur pFP39 (LFY) produit par F. Parcy dérive du vecteur p423 (expression inducible au galactose; prototrophie pour l'histidine; Mumberg et al., 1995). Le vecteur rapporteur pFP30 (AP1::lacZ) produit par F. Parcy dérive du vecteur pLG178 (prototrophie pour l'uracile; Guarente, 1983).

Pour l'étude de l'effet des mutations affectant la coopérativité, les vecteurs pETH31 (LFY FL H387A/R390A:VP16) et pETH32 (LFY FL H387A:VP16) produits par E. Thevenon et pFP14 (LFY FL:VP16) produit par F. Parcy dérivent du vecteur initial p424 (expression inducible au galactose; prototrophie pour le tryptophane; Mumberg et al., 1995). Le vecteur rapporteur pFP50 (AG::lacZ contenant une copie des sites de liaison AG-I et AG-II de LFY et WUS) produit par F. Parcy dérive du vecteur pLG178 (prototrophie pour l'uracile; Guarente, 1983).

Pour l'étude de la synergie d'activation de LFY et WUS, le vecteur pJL36 (WUS) produit par J. Lohmann (Institut Max Planck, Tuebingen, Allemagne) dérive du vecteur p423 (expression inducible au galactose; prototrophie pour l'histidine; Mumberg et al., 1995). Les vecteurs pFP13 (LFY), pFP14 (LFY:VP16) produits par F. Parcy et pETH9 (LFYΔ), pETH10 (LFYΔ:VP16), pETH31 (LFY FL H387A/R390A:VP16) et pETH34 (LFY FL H387A/R390A) produits par E. Thevenon dérivent du vecteur p424 (expression inducible au galactose; prototrophie pour le tryptophane; Mumberg et al., 1995). Le vecteur rapporteur pRH18 (AG::lacZ contenant trois copie des sites de liaison AG-I et AG-II de LFY et WUS pour accentuer la détection de la synergie d'activation) produit par R. Hong (Institut Max Planck, Tuebingen, Allemagne) dérive du vecteur pLG178 (prototrophie pour l'uracile; Guarente, 1983).

3. Expression et purification des protéines recombinantes

Le protocole de purification des protéines LFY-C sauvage et mutantes a été décrit dans Hamès et al., se référer à l'annexe 1.

Protocole d'induction alternatif des protéines LFY-C

Les vecteurs d'expression de LFY-C, et les versions mutantes R390A, H387A et double mutante R390A H387A ont été introduits dans des bactéries *E. coli* souche Rosetta2(DE3)pLysS (Novagen). Les bactéries sont ensemencées sous agitation continue le matin dans 5mL de milieu LB+antibiotiques (Kanamycine 50μg/mL + Chloramphenicol 34μg/mL) auxquels sont ajoutés à 2 reprises et à 45min d'intervalle 5mL de milieu LB+antibiotiques+NaCl 0,5M. Le tout est transféré en début d'après midi dans 1L de milieu LB+antibiotiques+NaCl 0,5M à 37°C jusqu'à atteindre une densité optique à 600nm (DO₆₀₀) de 0,600. Sont alors ajoutés 2mM de bêtaïne puis l'expression des protéines est induite sur 12 à 15h à 17°C par ajout de 0,4mM IPTG. Les bactéries induites sont centrifugées 30min à 4500g.

Protocole d'induction de LFY et WUS

Les vecteurs d'expression de LFY et WUS ont été introduits dans des bactéries *E. coli* souche RosettaBlue(DE3)pLysS (Novagen). A partir d'une préculture saturée, les bactéries sont diluées au 1/50^{ème} dans 1 à 8L de milieu TB+antibiotiques à 37°C en agitation continue. A DO₆₀₀ comprise entre 0.8 et 1, l'expression de LFY ou WUS est induite sur 12 à 15h à 22°C par ajout de 0,5mM IPTG. Les bactéries induites sont centrifugées 30min à 4500g.

Protocole de purification de LFY FL

LFY FL a été produit selon deux protocoles.

Protocole 1 suivi à partir des produits d'expression des vecteurs pCH2 (GST-Thrombine-LFY-6his) et pJUL1 (GST-Thrombine-LFY-GFP-6his)

Après la centrifugation de 30min à 4500g, chaque culot de 2 L de culture initiale est ressuspendu dans 35mL de tampon de lyse GST (20mM Tris pH8, 250mM NaCl, 1mM EDTA, 1mM azide Na, 5% [v/v] glycérol, 0,2% [v/v] triton X-100, 2mM DTT, une tablette de cocktail d'inhibiteur de protéases (Roche)), soniqué, et centrifugé 40min à 30000g. Le surnageant correspond à la fraction des protéines solubles.

Une purification reposant sur la présence de l'étiquette GST est réalisée par chromatographie d'affinité sur résine Glutathione sepharose 4B (Amersham BioSciences). Les protéines GST-LFY sont fixées à la résine par incubation 1h à 4°C (à raison de 2 mL de résine/L de culture initiale), puis centrifugées à 300g pendant 5min. La fraction non retenue est conservée comme contrôle. La résine est déposée sur colonne et lavée avec 10 volumes de TBSE (20mM Tris pH8, 250mM NaCl, 1mM EDTA, 5% [v/v] glycérol, 2mM DTT). Les protéines recombinantes sont éluées avec du tampon TBSE + glutathione réduite 100mM pH7, ou digérées pendant 2h (temps optimal de la digestion) à la thrombine (à raison de 25 unités/ml de résine) qui reconnaît un site de

clivage spécifique entre la GST et LFY. Dans ce dernier cas, la thrombine est inactivée avec du PMSF et éliminée avec des billes de p-aminobenzamidine. Les fractions obtenues sont analysées par électrophorèse sur gel SDS-page pour repérer la protéine d'intérêt et évaluer le rendement de la purification.

Une deuxième étape de purification basée sur la présence de l'étiquette 6his est réalisée par chromatographie d'affinité sur résine Ni sepharose (Amersham BioSciences). Les fractions contenant les protéines GST-LFY-6His ou LFY-6His sont dialysées une nuit à 4°C dans 2L de tampon de charge histidine (20mM Tris pH8, 500mM NaCl, 5% [v/v] glycérol, 1mM DTT, 5mM imidazole), incubées 1h à 4°C avec des billes Ni-NTA préalablement lavées et équilibrées (à raison de 250µL de résine Ni-NTA/L de culture initiale), puis centrifugées à 300g pendant 5min. Après dépôt sur colonne, les billes sont lavées à 5mM et 60mM en imidazole (concentration optimisée par gradient). L'élution est réalisée à 300mM en imidazole, les fractions obtenues sont déposées sur gel SDS-page.

Protocole 2 suivi à partir des produits d'expression du vecteur pCH52 (6his-GST-TEV-LFY)

Les produits d'expression obtenus à partir d'un litre de culture de pCH52 sont soumis au même protocole de purification que les produits d'expression de pCH28 (se reporter à l'annexe 1, Hamès et al.). Les protéines LFY ne sont donc purifiées que sur la présence de leur étiquette 6his. Des essais de clivage de la double étiquette 6his-GST par digestion TEV à 4°C pendant la nuit ont été entrepris à raison de 1% [m/m] de TEV par rapport à la masse totale en protéine. Ces essais ont été suivi d'une chromatographie d'affinité pour l'étiquette 6his qui permet de séparer les protéines LFY clivées des double étiquettes 6his-GST et de la TEV.

Protocole de purification de WUS FL et WUS Hd

La protéine WUS FL et le sous-domaine WUS Hd ont été purifiés selon le même protocole que LFY-C (décrit dans l'article Hamès et al., annexe 1) sur le critère de leur étiquette 6his.

Protocole de purification de la protéase TEV

Les vecteurs His TEV et His TEV point mutation ont été purifiées selon le protocole décrit dans (Parks et al., 1995). De meilleurs rendements ont été obtenus à partir du vecteur TEV His selon le protocole d'E. Derivery (Institut Curie, Paris; communications personnelles). Pour cela, les bactéries sontensemencées pendant la durée de la nuit dans une préculture de 25mL de milieu TB + kanamycine (50µg/mL) + chloramphenicol (35µg/mL). Le lendemain matin, le tout est transféré dans 1L de TB + antibiotiques et lorsque la DO₆₀₀ atteint 0.8, l'expression de la TEV est induite pendant 3h à 30°C par ajout de 1mM IPTG. Les bactéries sont ensuite centrifugées 30min à 4500g. Chaque culot de 1L de culture initiale est ressuspendu dans 2mL de solution PBS 2X, 0.6M NaCl (soit 0.45M NaCl sachant que le PBS contient déjà 0.15M), 1% Triton X100, 40mM imidazole, une tablette de cocktail d'inhibiteur de protéases (Roche) et mis à agiter doucement 3min à 30°C. Sont ensuite rajoutés 25mL de solution PBS 1X, NaCl 0.3M, Triton X100 0.1%, Imidazole 20mM, inhibiteur de protéases. Après 45min d'agitation douce à 4°C, le mélange est centrifugé 30min à 30000rpm. Le surnageant contenant les protéines solubles est mis en contact de billes Ni sepharose (Amersham BioSciences) sous agitation douce 2h à 4°C. Les protéines non hybridées à la résine sont lavées avec une solution 50mM Tris pH8, 0.1% Triton X100, 20mM imidazole. Les protéines hybridées sont elles éluées à l'aide d'une solution 50mM Tris pH8, 0.1% Triton X100, 200mM imidazole, puis dialysées pendant la nuit dans une solution 50mM Tris pH8, 1mM EDTA, 2mM DTT, 50% glycérol.

4. Etude biochimique des protéines

a) Gel d'acrylamide

Deux types de gels d'acrylamide sont réalisés: des gels en conditions dénaturantes (SDS-PAGE) pour évaluer la qualité et la quantité des protéines au cours des différentes étapes de purification, et des gels en conditions non dénaturantes ('natifs') pour les tests de liaison de LFY à l'ADN.

-SDS-PAGE

Gel de concentration: 5% acrylamide/bisacrylamide 37.5:1 (Merck), 62.5mM Tris-HCl pH6.8, 0.1% SDS, 0.1% persulfate d'ammonium (PSA), 0.01% TEMED

Gel de séparation: 10.5 ou 15% acrylamide/bisacrylamide 37.5:1 (Merck), 375mM Tris-HCl, 0.1% SDS, 0.1% PSA, 0.01% TEMED

Remarque: les concentrations en acrylamide/bisacrylamide sont ajustées en fonction de la masse moléculaire de la protéine à observer: 10,5% pour le suivi des étapes de purification des protéines LFY et WUS

FL (de masse moléculaire supérieure à 45kDa) ou 15% pour le suivi des étapes de purification des protéines LFY-C et WUS Hd (de masse moléculaire de l'ordre de 25kDa)

-natif 6% (pour les expériences d'EMSA)

6% acrylamide/bisacrylamide 29:1 (Sigma), 0.5X TBE, 0.1% PSA, 0.01% TEMED

b) Western Blot contre la protéine WUS

Protocole

Les protéines d'intérêt sont déposées sur un gel de polyacrylamide SDS-PAGE. Après 1h15 de migration à 160V, le gel est équilibré dans 20mL de tampon de transfert afin d'éliminer le SDS en excès. Les protéines sont transférées du gel de polyacrylamide à une membrane de nitrocellulose (Whatman) par migration dans du tampon de transfert à 100V pendant 1h et à 4°C. Suite au transfert, la membrane de nitrocellulose est équilibrée pendant 15min dans du tampon TBS_i à température ambiante et sous agitation continue puis saturée à deux reprises dans du tampon TBS_i-lait, la première fois pendant 5min, la seconde pendant une nuit à 4°C (ou 1h à température ambiante). La membrane est ensuite mise en présence des anticorps primaire anti-WUS (gracieusement données par J. Lohmann, Institut Max Planck, Tuebingen, Allemagne) dilués au 1/5000^{ème} dans 25mL de tampon TBS_i-lait et mise à incuber 2h à température ambiante (ou une nuit à 4°C). Trois lavages successifs de 5min dans du tampon TBS_i-lait permettent d'éliminer les anticorps non fixés. La membrane est ensuite exposée aux anticorps secondaires anti-lapin conjugués HRP (peroxydase de raifort, Jackson ImmunoResearch) dilués au 1/5000^{ème} dans du tampon TBS_i-lait pendant 30min à 1h à température ambiante. La membrane est lavée 5min dans du tampon TBS_i-lait puis à 3 reprises dans du tampon TBS_i pour éliminer toute trace d'anticorps secondaires anti-lapin conjugués HRP non fixés à la membrane. L'hybridation spécifique de ces derniers sur la membrane de nitrocellulose est révélée grâce à une réaction de chimioluminescence à l'aide du kit ECLplus (GE Healthcare) par mélange pendant 1h (maximum) à température ambiante des réactifs 1 et 2. La membrane est enfin exposée en cassette avec une membrane Amersham Hyperfilm ECL (GE Healthcare) pendant 5min, transférée sous lumière rouge dans du révélateur (Kodak) jusqu'à temps que les bandes d'intérêt soit suffisamment visibles, rincée dans de l'eau, transférée dans un fixateur (Kodak) et finalement rincée une dernière fois dans de l'eau.

Tampons

Tampon de transfert 1X: 25mM Tris pH8.3, 250mM glycine, 15% éthanol

Tampon TBS_i: 20mM Tris pH 7.4, 1X TBS, 150mM NaCl, 0.1% Triton X-100

Tampon TBS_i-lait: Tampon TBS_i, 5% [m/v] de lait écrémé en poudre (Régilait)

Tampons ECL:

-Solution A: 90mM acide p-coumarique (Sigma; dissout dans du DMSO)

-Solution B: 250mM luminol (Fluka; dissout dans du DMSO)

-Réactif 1 (quantité pour une réaction): 10mL Tris 100mM pH8.5, 13,3µL solution A, 1,66µL peroxyde d'hydrogène

-Réactif 2 (quantité pour une réaction): 10mL Tris 100mM pH8.5, 100µL solution B

c) Dosage de protéines

Dosage de Bradford

La protéine pure issue de pCH28 destinée à l'étude biochimique a été quantifiée par dosage au réactif de Bradford (BIO RAD): les valeurs de DO₅₉₅ de trois dilutions de la protéine sont confrontées à une gamme étalon de BSA titrée réalisée dans les mêmes conditions de mesures.

Mesure de l'absorption à 280nm

La protéine pure issue de pCH28 destinée à l'étude cristallographique a été quantifiée par mesure d'absorption à DO₂₈₀. D'après la loi de Beer-Lambert, $A = \epsilon Cl$, avec A = absorption à 280nm, ϵ = indice d'absorption, C = concentration, et l = longueur du trajet optique. La concentration en LFY correspond donc au quotient de cette mesure de DO par l'indice spécifique d'absorption à 280nm de LFY-C égale à 1,25.

Dosage sur gel d'acrylamide SDS-PAGE

Cette technique de dosage consiste à déposer un volume croissant de protéines sur gel SDS-PAGE et comparer l'intensité de la bande correspondante à la protéine d'intérêt à celle d'une gamme étalon d'actine titrée mise à migrer dans les mêmes conditions. Ces intensités sont estimées à l'aide du logiciel NIH. Cette technique a été sollicitée pour le dosage de protéines impures où seule la bande correspondante à la protéine d'intérêt a été quantifiée.

d) Anisotropie de fluorescence (AF) et expériences de retard sur gel (EMSA)

Oligonucléotides double brin cibles de LFY

Les oligonucléotides AP1 sauvage et mutants ont été dessinés d'après Parcy et al. (1998), et les oligonucléotides AG d'après Busch et al. (1999). Les oligonucléotides sont ressuspendus dans un tampon 1mM Tris pH7.5. Les oligonucléotides sens marqués en 5' par une molécule de TAMRA (pic d'excitation à 546nm, pic d'émission à 576nm, Sigma-Genosys), sont hybridés avec des oligonucléotides antisens complémentaire non marqués dans un tampon d'hybridation 1X (150mM NaCl, 1mM EDTA, 10mM Tris HCl pH7.5) par passage à 95°C pendant 2 à 3min et refroidissement progressif jusqu'à température ambiante.

Technique d'AF (Anisotropie de Fluorescence)

Initialement les données d'AF ont été obtenues à partir du fluorimètre MOS-450 (BioLogic Science Instruments) où les échantillons analysés (ADN + protéines + tampon de liaison ajusté à 50 µl final) étaient transférés dans des cuves en quartz. Le laboratoire s'étant équipé d'un Tecan Safire² (MTX Lab Systems) au cours de ma thèse, j'ai pu procéder dès lors à l'analyse d'échantillons de 40µl final en plaques de 384 puits. Dans le cas du fluorimètre MOS-450, les échantillons sont excités par un laser émettant une lumière à 546nm polarisée alternativement selon un axe vertical et horizontal à l'aide d'un modulateur de fréquence 100 kHz. Un photomultiplicateur avec filtre *cut off* centré à 578nm, placé à 90° de la lumière d'excitation, permet de convertir la fluorescence de l'échantillon mesurée en signal électrique amplifié. Dans le cas du Tecan Safire, les échantillons sont excités par une diode électroluminescente (DEL) produisant un rayonnement à 530nm. Cette lumière est ensuite polarisée et acheminée au monochromateur final par l'intermédiaire d'un système mettant en jeu trois lentilles optiques: la première lentille collecte et aligne selon un axe vertical la lumière produite par la DEL; la deuxième focalise cette lumière d'excitation polarisée sur les puits puis collecte la lumière plus ou moins polarisée émise par l'échantillon; la troisième focalise la lumière d'émission sur une fibre optique qui transmet le signal lumineux à un monochromateur centré sur 580nm. Un photomultiplicateur convertit en dernier lieu le signal lumineux en signal électrique amplifié.

Le calcul d'AF est effectué comme suit (Canet et al., 2001): $AF = (V_v - V_h) / (V_v + 0,5V_h)$, où V_v est la lumière émise après excitation selon l'axe vertical, et V_h est la lumière émise après excitation selon l'axe horizontal.

Ces données permettent de déduire le K_d apparent de la réaction de liaison (en supposant qu'une molécule de LFY se lie à une molécule d'ADN).

En effet, $AF_{\text{mesurée}} = A_f F_f + A_b F_b$ avec

- A_f : AF des oligonucléotides marqués libres
- F_f : fraction des oligonucléotides marqués libres
- A_b : AF des oligonucléotides marqués liés
- F_b : fraction des oligonucléotides marqués liés

$$\text{Sachant } F_f + F_b = 1 \text{ et } F_b = \frac{[ADN-LFY]}{[ADN_{\text{total}}]}, \text{ alors } AF_{\text{mesurée}} = A_f + (A_b - A_f) \frac{[ADN-LFY]}{[ADN_{\text{total}}]} \quad (1)$$

En considérant la loi d'action de masse définissant la constante de dissociation (K_d)

$$\text{dans la réaction } [ADN] + [LFY] \rightleftharpoons [ADN-LFY]: K_d = \frac{[ADN][LFY]}{[ADN-LFY]}$$

Et la loi de conservation des éléments: $[ADN_{\text{total}}] = [ADN] + [ADN-LFY]$ et $[LFY_{\text{total}}] = [LFY] + [ADN-LFY]$

$$\text{Nous pouvons écrire: } [ADN-LFY]^2 + [ADN-LFY](-K_d - [ADN_{\text{total}}] - [LFY_{\text{total}}]) + [ADN_{\text{total}}][LFY_{\text{total}}] = 0 \quad (2)$$

D'après (1) et par résolution de (2):

$$A = A_f + (A_b - A_f) \frac{(K_d + [ADN_{\text{total}}] + [LFY_{\text{total}}]) - \sqrt{(K_d + [ADN_{\text{total}}] + [LFY_{\text{total}}])^2 - 4[ADN_{\text{total}}][LFY_{\text{total}}]}}{2[ADN_{\text{total}}]}$$

Technique d'EMSA (Electrophoretic Mobility Shift Assay)

Ce protocole a été décrit dans l'article Hamès et al., en annexe 1, ainsi que les calculs de K_{d1} et K_{d2} .

La loi de conservation des éléments et la loi d'action de masse nécessaires à l'estimation du K_d ont été décrites dans le paragraphe précédent.

Cas de l'EMSA en gel gradient linéaire à l'équilibre (GEMSA):

Le gradient linéaire a été réalisé à l'aide d'un système de formeur de gradient à deux chambres communicantes, une chambre comprenant la solution d'acrylamide la plus concentrée (29%), l'autre la moins concentrée (5%). La chambre contenant la solution la plus concentrée est reliée à une pompe péristaltique qui

déverse le gradient en cours de formation entre les plaques de verre cloisonnant le futur gel. Le gel final est donc très concentré en acrylamide à sa base, moins concentré au niveau des puits.

La réaction de liaison est réalisée dans un volume final de 50µl de tampon de liaison (150 mM NaCl, 20 mM Tris-HCl pH 7.5, 1 % glycerol), supplémenté de 28 ng/µl d'ADN de sperme de poisson (Roche) et 10nM d'oligonucléotides AP1 double brin.

La migration se fait à 400V (soit environ 15mA au démarrage) pendant environ 20h (jusqu'à ce que l'ampérage atteigne une valeur quasi-nulle signalant l'atteinte de l'équilibre de migration) et à 4°C dans un tampon 0,5X TBE.

La droite de calibration est obtenue en représentant le logarithme de la masse moléculaire en fonction de la distance de migration normalisée par rapport au front de migration total. La masse moléculaire des différents complexes (MW, *Molecular Weight*) est alors égale à: $MW=10^{[ax+b]}$ où x correspond à la distance de migration normalisée de la bande d'intérêt (Fig. A).

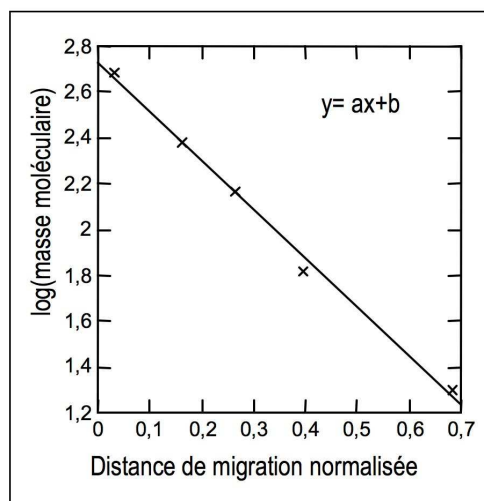


Figure A: Méthode de calibration des gels GEMSA

Représentation graphique de la droite de référence $\log(MW) = f(\text{distance de migration de la protéine standard rapportée au front de migration})$. Données recueillies avec Bande2-Apoferritine (480kDa), B-phycoerythrine (242kDa), Lactate déhydrogenase (146kDa), Albumine de sérum bovin (ASB) (66kDa) et Inhibiteur de trypsine de soja (20kDa).

Solution d'acrylamide concentré à 29% (à raison de 11,5mL total par gel, spacers 0,75mm): 29% acrylamide/bisacrylamide 29:1 (Sigma), 0,5X TBE, 13% saccharose (soit 1,5g par gel), 0,1% PSA, 0,01% TEMED

Solution d'acrylamide diluée à 5% (à raison de 11,5mL totale par gel, spacers 0,75mm): 5% acrylamide/bisacrylamide 29:1 (Sigma), 0,5X TBE, 13% saccharose (soit 1,5g par gel), 0,1% PSA, 0,01% TEMED

Cas du Ferguson plot (Orchard and May, 1993; York and Reznikoff, 1996)

La réaction est réalisée selon le même protocole que les EMSA classiques, à partir de gels non détergents à concentrations allant de 5 à 11% d'acrylamide. La migration se fait à 70V pendant 15min puis à 90V pendant 1h05 pour des gels à 5% d'acrylamide, à 90V pendant 1h25 pour des gels à 6%, à 90V pendant 1h45 pour des gels à 7 et 8%, et à 90V pendant 2h20 pour des gels de 9, 10 ou 11% d'acrylamide.

La calibration de cette méthode est expliquée en figure B. La masse moléculaire du complexe principal est déduite de la comparaison de sa distance de migration à celles des protéines standard à partir de l'équation de la droite de référence $y = ax + b$ où $y = \log(\text{pente})$ et $x = (\log MW)$ (Fig.B; kit NativeMark Invitrogen), tel que $MW=10^{[\log(\text{pente})/a]}$.

e) Technique de SEC (*Size Exclusion Chromatography*) analytique

Le protocole de SEC analytique a été décrit dans Hamès et al. (annexe 1). La calibration a été réalisée dans les mêmes conditions de tampon à partir du kit Gel filtration calibration (GE Healthcare; thyroglobuline 669kDa, Ferritine 440kDa, Catalase 232kDa, Albumine de sérum bovin 66kDa, Ribonuclease A 13kDa; Fig.C). La masse moléculaire de LFY-C, AP1, et LFY-C/AP1 est déduite de l'équation de référence $y = be^{ax}$ où $y = MW$ et $x = \text{volume d'élution}$.

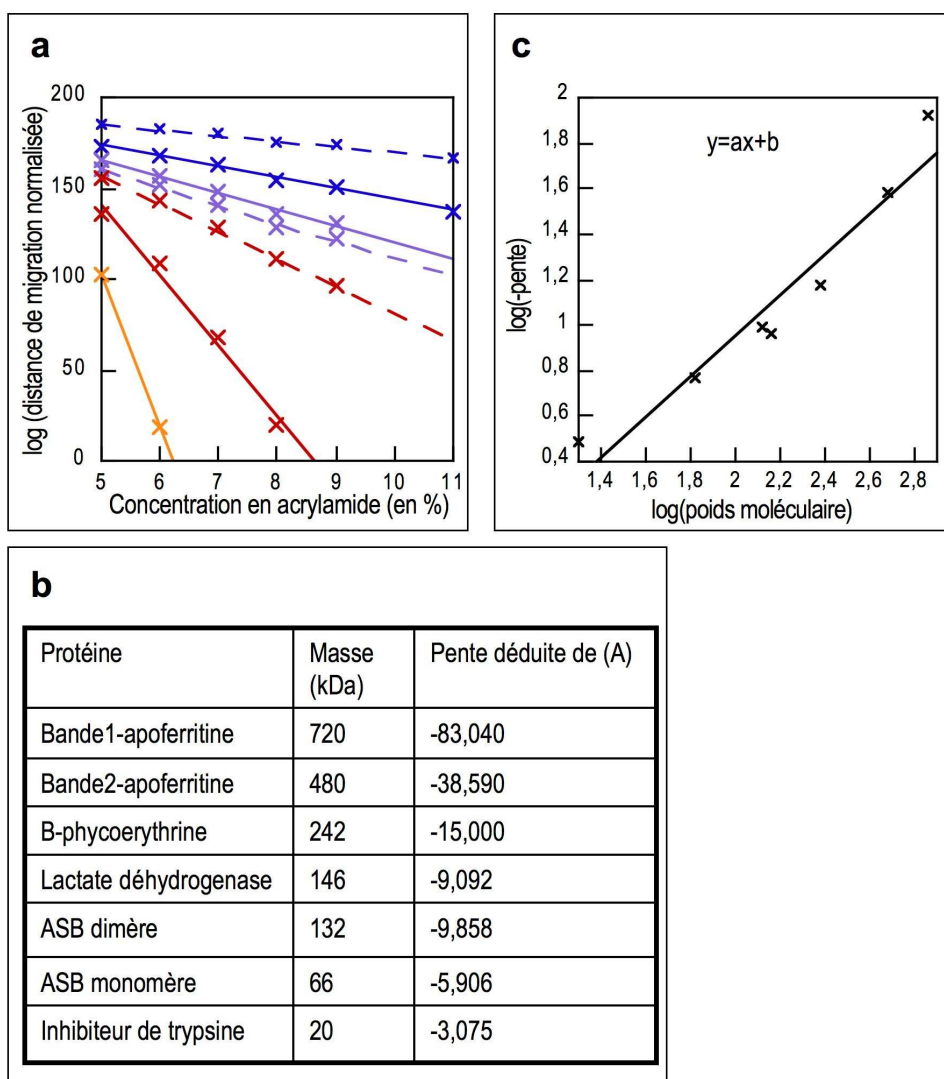


Figure B: Méthode de calibration de la technique de Ferguson Plot

(a) Représentation graphique de $\log(\text{distance de migration de la protéine standard rapportée au front de migration}) = f(\% \text{ d'acrylamide})$; Bande1-apoferritine 720kDa en orange, Bande2-apoferritine 480kDa en rouge; B-phycoerythrine 242kDa en pointillés rouges; Lactate déshydrogénase 146kDa en violet; Albumine de sérum bovin (ASB) Dimère 132kDa en pointillés violets; ASB Monomère 66kDa en bleu; Inhibiteur de trypsine de soja 20kDa en pointillés bleus; (b) Tableau récapitulatif des pentes obtenues pour chacune de protéines marqueurs; (c) Représentation graphique de la droite de référence $\log(-\text{pente}) = f(\log(\text{poids moléculaire}))$

f) Technique de MALLS-SEC (Multi-Angle Laser Light Scattering)

La technique de MALLS-SEC a été décrite dans Gérard et al., 2007 (Gerard et al., 2007).

g) Technique de SPR (Surface Plasmon Resonance)

La technique de SPR a été décrite dans Moyroud et al., annexe 2.

5. Cristallogénèse

a) LFY-C/ADN

Les premiers essais réalisés à l'aide du robot de l'EMBL à température ambiante ont été effectués à partir d'une protéine 6his-LFY-C à 6mg/mL et selon un ratio molaire 1:4 AP1:protéine.

Les essais suivants ont tous été réalisés manuellement à 4°C en système de diffusion de vapeur en goutte suspendue. Les protéines ont été débarrassées de leur étiquette 6his par clivage à la protéase TEV (cf protocole de purification). A partir d'une protéine concentrée à 5mg/mL, nous avons pratiqué un ratio molaire 1:4 protéine:ADN dans un premier temps avant de le rectifier en un ratio molaire 1:2. Les différents mélanges de protéines issues du vecteur pCH26 ou pCH28 avec les oligonucléotides AP1, AG-I et AG-II de taille variable

(tableau 2) ont été testés dans les 48 conditions du kit Natrix (Hampton Research) à raison de 1µL de mélange protéine/ADN + 1µL de tampon de tampon Natrix par goutte. Le tableau B résume toutes les conditions ayant conduit à la formation de cristaux, qu'ils aient été ou non analysés au synchrotron.

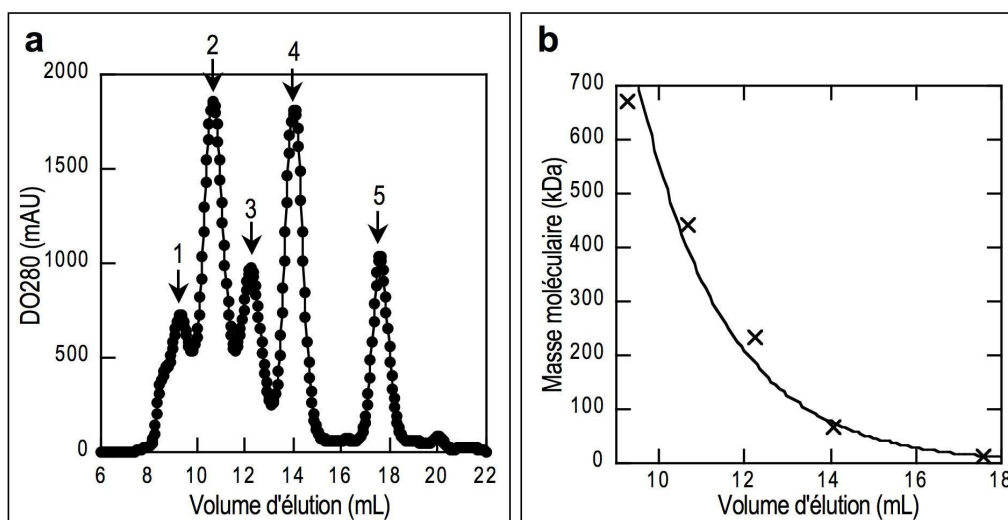


Figure C: Définition de la gamme de protéines standard par SEC analytique

(a) Profil de SEC analytique (DO_{280}) des protéines standard. Les flèches signalent l'élution de 1) la thyroglobuline 669kDa, 2) la ferritine 440kDa, 3) la catalase 232kDa, 4) l'albumine de sérum bovin 66kDa et 5) la ribonucléase a 13kDa; (b) Représentation graphique du volume d'élution des protéines standard en fonction de leur masse moléculaire. Les données expérimentales sont ajustées à une courbe logarithmique permettant d'établir l'équation de référence $y = be^{ax}$

b) LFY-C

Des essais de cristallogénèse de la protéine LFY-C seule ont été réalisés à des concentrations de 2,5mg/mL et 5mg/mL de protéine et dans les 48 conditions des kits Crystal Screen et Crystal Screen Lite (Hampton research). Aucune condition n'a permis d'obtenir des cristaux de protéines.

6. Caractérisation de LFY chez la levure

a) Milieux de culture

Les levures poussent comme les bactéries en milieu liquide ou solide, mais à 30°C, et en général pendant 2 jours. La sélection se fait ici en fonction de leur capacité à croître sur un milieu appauvri : le vecteur qu'elles reçoivent leur permet en général de produire un acide aminé, du tryptophane par exemple, dont la présence n'est donc plus nécessaire dans le milieu de culture.

Milieu complet YPD: 1% [m/v] yeast extract, 2% [m/v] peptone, 2% [m/v] difco bacto agar, 2% [m/v] glucose filtré
SM (milieu minimum): 0.17% [m/v] YNB, 0.5% [m/v] NH_4SO_4 , poudre AA (fonction de la prototrophie), 2% [m/v] glucose ou galactose filtré + optionnel: 2% agar pour obtenir un milieu solide.

Remarque : dans les cas présentés ici le milieu minimum est UTH- (sans uracile, tryptophane et histidine). On rajoute alors 0,70g/L de poudre UTH- (BIO 101 Systems, Q-BIOgene).

b) Transformation des levures

Le protocole de transformation des levures est accessible sur le site <http://www.umanitoba.ca/faculties/medicine/units/biochem/gietz/trafo.html> (Gietz et al., 1995).

Les levures sont mises en culture dans 2mL de milieu SM jusqu'à saturation. 20µL sont ensuite transférés dans 25mL de milieu SM. A DO_{600} égale à 1, les levures sont centrifugées pendant 5min à 3000g puis ressuspendues dans 50mL de milieu préchauffé SM (ou YPD). A DO_{600} égale à 0,6-0,7, les levures sont de nouveau centrifugées puis ressuspendues dans 15mL d'eau stérile. 3mL sont répartis dans 5 tubes eppendorf à raison de 2X1,5mL par tube, avec deux centrifugations à 6000rpm. Sont rajoutées successivement au culot de levures 480µL de PEG 3500 (polyéthylène glycol) 50%, 72µL d'acétate de lithium 1M pH7.5, 50µL d'ADN simple brin, et 100µL d'eau stérile contenant entre 50ng et 5µg de plasmide. Le culot est ensuite vortexé vigoureusement jusqu'à son entière dissolution, puis incubé 30min à 30°C. Les levures subissent finalement un choc thermique de 15min à 42°C avant d'être centrifugées, ressuspendues dans 500µL d'eau stérile, et incubées sur boîte SM+agar pendant 2 à 4 jours.

Tableau B: Récapitulatif des différentes conditions de cristallogenèse ayant abouti à la formation de cristaux de protéines

Les protéines issues du vecteur pCH26 et pCH28 sont respectivement dénommées ici C26 et C28.

Mélange protéine/ADN	N° Natrix	Composition du tampon
C26/AG-I 21pb	37	0.2M Potassium chloride, 0.025M Magnesium sulfate hydrate, 0.05M HEPES sodium pH7.0, 20% v/v PEG 200
	38	0.2M Ammonium acetate, 0.15M Magnesium acetate tetrahydrate, 0.05M HEPES sodium pH7.0, 5% m/v PEG 4000
	39	0.1M Ammonium acetate, 0.02M Magnesium chloride hexahydrate, 0.05M HEPES sodium pH7.0, 5% m/v PEG 8000
	47	0.1M Potassium chloride, 0.01M Magnesium chloride hexahydrate, 0.05M Tris hydrochloride pH8.5, 30% v/v PEG 400
C26/AP1 symétrique 20pb	3	0.01M Magnesium acetate tetrahydrate, 0.05M MES monohydrate pH5.6, 20% v/v (+/-)-2-Methyl-2,4-pentanediol
	4	0.2M Potassium chloride, 0.01M Magnesium sulfate heptahydrate, 0.05M MES monohydrate pH5.6, 10% v/v PEG 400
	5	0.2M Potassium chloride, 0.01M Magnesium chloride hexahydrate, 0.05M MES monohydrate pH5.6, 5% m/v PEG 8000
	6	0.1M Ammonium sulfate, 0.01M Magnesium chloride hexahydrate, 0.05M MES monohydrate pH5.6, 20% m/v PEG 8000
	7	0.02M Magnesium chloride hexahydrate, 0.05M MES monohydrate pH6.0, 15% v/v 2-Propanol
	8	0.1M Ammonium acetate, 0.005M Magnesium sulfate heptahydrate, 0.05M MES monohydrate pH6.0, 0.6M Sodium chloride
	9	0.1M Potassium chloride, 0.01M Magnesium chloride hexahydrate, 0.05M MES monohydrate pH6.0, 10% v/v PEG 400
	10	0.005M Magnesium sulfate heptahydrate, 0.05M MES monohydrate pH6.0, 5% m/v PEG 4000
	15	0.04M Magnesium chloride hexahydrate, 0.05M Sodium cacodylate trihydrate pH6.0, 5% v/v (+/-)-2-Methyl-2,4-pentanediol
	17	0.2M Potassium chloride, 0.01M Calcium chloride dihydrate, 0.05M Sodium cacodylate trihydrate pH6.0, 10% m/v PEG 4000
	20	0.1M Ammonium acetate, 0.015M Magnesium acetate tetrahydrate, 0.05M Sodium cacodylate trihydrate pH6.5, 10% v/v 2-Propanol
	22	0.08M Magnesium acetate tetrahydrate, 0.05M Sodium cacodylate trihydrate pH6.5, 15% v/v PEG 400
	23	0.2M Potassium chloride, 0.01M Magnesium chloride hexahydrate, 0.05M Sodium cacodylate trihydrate pH6.5, 10% m/v PEG 4000
	24	0.2M Ammonium acetate, 0.01M Calcium chloride dihydrate, 0.05M Sodium cacodylate trihydrate pH6.5, 10% m/v PEG 4000
	31	0.005M Magnesium chloride hexahydrate, 0.05M HEPES sodium pH7.0, 25% PEG monomethyl ether 550
	39	0.1M Ammonium acetate, 0.02M Magnesium chloride hexahydrate, 0.05M HEPES sodium pH7.0, 5% m/v PEG 8000
	41	0.1M Potassium chloride, 0.015M Magnesium chloride hexahydrate, 0.05M Tris hydrochloride pH7.5, 10% v/v PEG monomethyl ether 550
	42	0.01M Magnesium chloride hexahydrate, 0.05M Tris hydrochloride pH7.5, 1% v/v 2-Propanol
	47	0.1M Potassium chloride, 0.01M Magnesium chloride hexahydrate, 0.05M Tris hydrochloride pH8.5, 30% v/v PEG 400
C26/AP1 18pb	3	0.01M Magnesium acetate tetrahydrate, 0.05M MES monohydrate pH5.6, 20% v/v (+/-)-2-Methyl-2,4-pentanediol
	5	0.2M Potassium chloride, 0.01M Magnesium chloride hexahydrate, 0.05M MES monohydrate pH5.6, 5% m/v PEG 8000
	9	0.1M Potassium chloride, 0.01M Magnesium chloride hexahydrate, 0.05M MES monohydrate pH6.0, 10% v/v PEG 400
	10	0.005M Magnesium sulfate heptahydrate, 0.05M MES monohydrate pH6.0, 5% m/v PEG 4000
	15	0.04M Magnesium chloride hexahydrate, 0.05M Sodium cacodylate trihydrate pH6.0, 5% v/v (+/-)-2-Methyl-2,4-pentanediol
	17	0.2M Potassium chloride, 0.01M Calcium chloride dihydrate, 0.05M Sodium cacodylate trihydrate pH6.0, 10% m/v PEG 4000
	20	0.1M Ammonium acetate, 0.015M Magnesium acetate tetrahydrate, 0.05M Sodium cacodylate trihydrate pH6.5, 10% v/v 2-Propanol
	23	0.2M Potassium chloride, 0.01M Magnesium chloride hexahydrate, 0.05M Sodium cacodylate trihydrate pH6.5, 10% m/v PEG 4000
	24	0.2M Ammonium acetate, 0.01M Calcium chloride dihydrate, 0.05M Sodium cacodylate trihydrate pH6.5, 10% m/v PEG 4000
	26	0.2M Potassium chloride, 0.1M Magnesium acetate tetrahydrate, 0.05M Sodium cacodylate trihydrate pH6.5, 10% m/v PEG 8000
	31	0.005M Magnesium chloride hexahydrate, 0.05M HEPES sodium pH7.0, 25% v/v PEG monomethyl ether 550
	36	0.1M Potassium chloride, 0.01M Calcium chloride dihydrate, 0.05M HEPES sodium pH7.0, 10% v/v PEG 400
	38	0.2M Ammonium acetate, 0.15M Magnesium acetate tetrahydrate, 0.05M HEPES sodium pH7.0, 5% m/v PEG 4000
	39	0.1M Ammonium acetate, 0.02M Magnesium chloride hexahydrate, 0.05M HEPES sodium pH7.0, 5% m/v PEG 8000
	41	0.1M Potassium chloride, 0.015M Magnesium chloride hexahydrate, 0.05M Tris hydrochloride pH7.5, 10% v/v PEG monomethyl ether 550
	47	0.1M Potassium chloride, 0.01M Magnesium chloride hexahydrate, 0.05M Tris hydrochloride pH8.5, 30% v/v PEG 400
C26/AG-II 21pb	38	0.2M Ammonium acetate, 0.15M Magnesium acetate tetrahydrate, 0.05M HEPES sodium pH7.0, 5% m/v PEG 4000
	44	0.2M Potassium chloride, 0.05M Magnesium chloride hexahydrate, 0.05M Tris hydrochloride pH7.5, 10% m/v PEG 4000
C26/AP1 19pb	33	0.2M Ammonium chloride, 0.01M Magnesium chloride hexahydrate, 0.05M HEPES sodium pH7.0, 30% m/v 1,6-Hexanediol
	36	0.1M Potassium chloride, 0.01M Calcium chloride dihydrate, 0.05M HEPES sodium pH7.0, 10% v/v PEG 400
	38	0.2M Ammonium acetate, 0.15M Magnesium acetate tetrahydrate, 0.05M HEPES sodium pH7.0, 5% m/v PEG 4000
	43	0.05M Ammonium acetate, 0.01M Magnesium chloride hexahydrate, 0.05M Tris hydrochloride pH7.5, 10% v/v (+/-)-2-Methyl-2,4-pentanediol
	45	0.025M Magnesium sulfate hydrate, 0.05M Tris hydrochloride pH8.5, 1.8M Ammonium sulfate
	46	0.005M Magnesium sulfate hydrate, 0.05M Tris hydrochloride pH8.5, 35% m/v 1,6-Hexanediol
C28/AP1 21pb	7	0.02M Magnesium chloride hexahydrate, 0.05M MES monohydrate pH6.0, 15% v/v 2-Propanol
	20	0.1M Ammonium acetate, 0.015M Magnesium acetate tetrahydrate, 0.05M Sodium cacodylate trihydrate pH6.5, 10% v/v 2-Propanol
	24	0.2M Ammonium acetate, 0.01M Calcium chloride dihydrate, 0.05M Sodium cacodylate trihydrate pH6.5, 10% m/v PEG 4000
C28/AP1 23pb	3	0.01M Magnesium acetate tetrahydrate, 0.05M MES monohydrate pH5.6, 20% v/v (+/-)-2-Methyl-2,4-pentanediol
	21	0.2M Potassium chloride, 0.005M Magnesium chloride hexahydrate, 0.05M Sodium cacodylate trihydrate pH6.5, 10% m/v 1,6-Hexanediol
	40	0.01M Magnesium chloride hexahydrate, 0.05M Tris hydrochloride pH7.5, 1.6M Ammonium sulfate

c) Protocole de mesure de l'activité β -galactosidase

-Analyse qualitative

Les colonies de levures sont repiquées d'une boîte maîtresse (milieu sélectif SM-UTH dans notre cas) SM 2% glucose sur une boîte SM 2% galactose. Après 48h minimum à 30°C, les levures sont transférées sur une membrane de nitrocellulose, ce qui conduit à un aspect non homogène des colonies de levures (Exp.1, Fig.48C). La membrane de nitrocellulose peut éventuellement être placée directement sur la boîte SM 2% galactose afin de repiquer les levures directement sur la membrane (Exp.2, Fig.48C).

Le tout est mis à sécher sur papier buvard, trempé 5min dans l'azote liquide, séché sur papier buvard et trempé à deux nouvelles reprises dans l'azote liquide selon ce même protocole de façon à lyser les cellules.

La membrane est ensuite transférée à 37°C sur papier buvard imbibé de tampon Z (60mM Na₂HPO₄, 12H₂O; 40mM NaH₂PO₄, 2H₂O; 10mM KCl, 1mM MgSO₄, 7H₂O), β -mercaptoéthanol 50mM et 5-

bromo-4chloro-3-indolyl b-D-galactopyranoside (X Gal) à 1mg/mL, en environnement clos (dans une boîte de pétri par exemple) jusqu'à observation d'une coloration bleue, manifestation de l'hydrolyse du substrat X-Gal par la β -galactosidase.

-Analyse quantitative

Une colonie isolée est repiquée dans 2mL de milieu sélectif (SM UTH dans notre cas) 2% glucose pendant 15h minimum. 5 μ L de culture saturée sont inoculés dans 3mL de milieu sélectif SM 2% galactose, puis le tout est incubé sous agitation continue sur la nuit à 30°C jusqu'à une DO₆₀₀ de 0,2 à 0,6. 1mL de cette culture est prélevé et mis à centrifuger à 3000g pendant 5min. Le culot de levure est ressuspendu à deux reprises dans 1mL de milieu sélectif 2% galactose. Le tout est incubé sous agitation continue pendant 6h à 30°C. Les levures sont ensuite centrifugées à 3000g pendant 5min et ressuspendues dans 0,40mL de tampon Z froid (décrit dans le paragraphe précédent). Après une nouvelle centrifugation à 6000g pendant 5min à 4°C, le culot est ressuspendu dans 0,35mL de tampon Z froid. Les levures sont alors perméabilisées par ajout de 16 μ L de chloroforme + 16 μ L de SDS 0,1%, vortexées vigoureusement, et incubées 30min à 30°C. 232 μ L de la phase aqueuse sont transférés dans un tube eppendorf auquel sont ajoutés 40 μ L d'ONPG à 4mg/mL (stocké dans du KPO₄ 0,1M pH7.0). La DO₅₉₅ est mesurée, puis l'activité *lacZ* de catalyse de l'ONPG (incolore) en ONP (jaune) + galactose est suivie par mesure de la DO₄₂₀ en cinétique sur 1h.

L'activité β -galactosidase est alors estimée en 'Miller units' à partir du calcul suivant:

Miller Units = $1000 \times DO_{420} / ((\text{volume de cellules}) \times \text{temps de réaction en min} \times DO_{595})$

Soit Miller Units = $1000 \times \text{pente initiale de la réaction} / (\text{volume de cellule} \times DO_{595})$

7. Caractérisation de LFY chez les plantes

a) Stérilisation des graines

Les graines sont exposées 3min dans 70%EtOH + 0,05%SDS, 1min dans 95% EtOH, puis mises à sécher sur du papier buvard stérile sous hotte.

b) Conditions de culture des plantes

Les graines d'*A. thaliana* sont mises à stratifier 2 jours à 4°C puis transférées en salle de culture photopériode jours longs (25°C, 16h de lumière, 8h d'obscurité) sur support autoclavé 50% vermiculite exfoliée, 50% terreau.

Les graines de *Nicotiana benthamiana* sont soumises au même traitement sans période de stratification préalable.

c) Expression transitoire en feuilles de tabac *N.benthamiana*

Les plasmides sont introduits par électroporation dans des agrobactéries (*Agrobacterium tumefaciens* C58C1 pMP90) puis étalées sur boîte LB sélective à 28°C. A partir d'une colonie isolée repiquée dans 5mL de milieu LB + antibiotiques, 1mL de culture saturée est transféré dans 4 mL de milieu inducteur pH 5,6 (50,78 mM MES, 0,5% [v/v] glucose, 1,734 mM NaH₂PO₄, 0,2 mM acétosyringone, 5% de mix 20X-AB [373,9 mM NH₄Cl, 24,34 mM MgSO₄, 40,23 mM KCl, 1,36 mM CaCl₂, 0,18 mM FeSO₄.7H₂O]) et mis à pousser à 28°C pendant 6 h jusqu'à une DO₆₀₀ comprise entre 0,6 et 1 (Lavy et al., 2002). Les agrobactéries sont infiltrées sur des jeunes feuilles de tabacs âgés de 4 à 7 semaines par pression sur l'épiderme inférieur à l'aide d'une seringue.

d) Transformation stable d'*A. thaliana* par *A. tumefaciens*

Milieus de culture

YEP: 1% [m/v] yeast extract, 1% [m/v] bactopeptone, 0.5% [m/v] NaCl, pH7.0

1/2MS: 1% [m/v] saccharose, 0.22% [m/v] MS (Murashige et Skoog), 0.05% [m/v] MES, 1% [m/v] agarose pour obtenir un milieu solide

Protocole de transformation

Les plasmides sont introduits par électroporation dans des agrobactéries (*A. tumefaciens* C58C1 pMP90) puis étalées sur boîte LB sélective à 28°C. Une colonie isolée est repiquée dans 5mL de milieu LB + antibiotiques, puis transférée à saturation dans 200mL YEP + antibiotiques. A DO₆₀₀ comprise entre 0,8 et 1, la culture est centrifugée 10min à 4000rpm et à température ambiante puis le culot est ressuspendu dans 300mL de milieu 1/2MS auxquels est rajouté 150 μ L de Silweet L-77. Les inflorescences d'*A. thaliana* ayant effectué leur montaison 1 à 2 semaines auparavant (tiges mesurant entre 10 et 20cm) sont trempées pendant 1 à 2min dans le mélange d'agrobactéries+1/2MS+Silweet L-77. Les plantes sont ensuite incubées sous cloche pendant un à

deux jours. Approximativement 3 semaines plus tard (lorsque les premières siliques arrivent à déhiscence), les graines sont récoltées quotidiennement.

e) Microscopie confocale

Les feuilles de tabac infiltrées sont visualisées 4 ou 5 jours après infiltration. La zone infiltrée est montée dans l'eau, et observée au microscope confocal inversé TCS-SP2 (Leica Microsystem) à l'objectif à immersion X20 ou X40. La GFP et la YFP (protéine fluorescente jaune) sont excitées à 488 nm et l'émission est mesurée entre 500 et 555 nm. mRFP1 est excitée à 543 nm et la lumière d'émission est récupérée dans l'intervalle [575 nm: 635 nm].

f) Extraction de l'ADN génomique (ADNg) des plantes d'*A. thaliana*

L'équivalent de 0,5mm X 0,5mm de matériel végétal de préférence jeune (jeunes feuilles ou zones méristématiques) est écrasé à température ambiante pendant 15 sec à l'aide d'un pilon adapté aux tubes eppendorf 1,5mL. 200µL de tampon d'extraction (200mM Tris-HCl pH7.5; 250mM NaCl; 25mM EDTA; 0,5%SDS) sont ajoutés puis le mélange est vortexé 5sec. Après centrifugation à 10000g pendant 1min, 150µL de surnageant sont transférés dans un tube eppendorf neuf. 150µL d'isopropanol sont rajoutés et le tout est laissé à -20°C pendant 2min. Après centrifugation à 10000g pendant 5min, le culot est lavé (optionnel) avec 100µL d'EtOH 70% et centrifugé à 10000g pendant 5min. Finalement, le culot est séché au speed vacuum puis dissout dans 50µL de TE 1X.

g) Géotypage des plantes *Ify-12*

L'ADNg des plantes à tester est amplifié par PCR à partir des oligonucléotides oLFY1 (5'AAGCAGCCGTCTGCGGTGTCAGCAGCTGTT) et oLFY2 (5'CTGTCAATTTCCCAGCAAGACAC).

Mélange: 1.25µL dNTP 10mM, 2.5µL Tampon PCR MgCl₂ 10X, 0.5µL oLFY1 20mM, 0.5µL oLFY2 20mM, 5µL d'extrait d'ADNg, 15.05µL eau distillée, 0.2µL Q-bio taq polymerase (MP Biomedicals, Qiagen)

Programme: 94°C 4min, (94°C 30 sec, 57°C 30 sec, 72°C 30 sec) X 35 cycles

Les produits PCR sont ensuite digérés 3h à 60°C.

Mélange: 15µL de produit PCR, 2µL eau distillée, 2µL Tampon W (NEBuffer New England BioLabs), 1µL BstAPI (New England BioLabs)

Enfin, le géotype des plantes est obtenu par visualisation sur gel d'acrylamide TAE 0,5X, 3% agarose: la taille des fragments de restriction attendue est 370pb pour l'allèle *Ify-12*, 345pb pour l'allèle *LFY*.

I. Caractérisation biochimique du domaine conservé LFY-C

1. Introduction: pourquoi étudier le domaine C-terminal ?

Le domaine C-terminal de LFY défini du fait de sa conservation (résidus 230-391) est un petit domaine de 25kDa (soit 160aa) dont le coeur est très bien structuré d'après les logiciels de prédiction (Fig.18A). C'est également le domaine putatif de liaison de LFY à l'ADN: dans les figures supplémentaires de leur article (Fig.13), Maizel et collaborateurs délimitent par EMSA le domaine minimal de liaison de LFY à l'ADN aux acides aminés 230-421, l'acide aminé 230 coïncidant parfaitement avec le début du domaine conservé C-terminal. Etrangement, ils observent également qu'une délétion partielle dans le domaine conservé N-terminal (délétions des résidus 1-58) abolit totalement la liaison de LFY à l'ADN, alors que la délétion complète de ce domaine (résidus 1-130) ne semble pas perturber cette interaction. Une explication plausible serait que la délétion dans le domaine conservé N-terminal entraîne un mauvais repliement de la protéine et par conséquent son inactivité, suggérant que ce domaine n'interviendrait pas dans la liaison de LFY à l'ADN.

Réduire l'étude d'un facteur de transcription à son domaine de liaison à l'ADN (DBD, DNA-binding domain) est une approche couramment utilisée car ce sont souvent des domaines au repliement assez indépendant du reste de la protéine et qui permettent ainsi d'améliorer le rendement de la production protéique (Luscombe et al., 2000). Cette stratégie a d'ores et déjà permis d'obtenir de nombreuses structures cristallographiques de DBD liés à leurs séquences régulatrices. C'est pourquoi nous avons décidé de nous focaliser sur l'étude du domaine conservé C-terminal.

2. Production du domaine C-terminal de LFY: sélection du vecteur d'expression optimal

Dans le but de cristalliser le domaine conservé C-terminal, j'ai construit différents vecteurs d'expression et testé lequel aboutissait à la plus grande quantité de protéine pure et active. Pour cela, l'ADNc codant pour la région C-terminale a été introduit dans deux vecteurs différents: le vecteur pIVEX2.4d (ADNc-6his; Roche); et le vecteur pETM-11 (6his-TEV-ADNc; collection EMBL). Quatre constructions ont été testées en parallèle en jouant sur les extrémités 3' et 5' afin d'identifier celle qui conduit au meilleur rendement, ces extrémités étant prédites comme mal structurées (Fig.18A). Deux sont dessinées au ras de l'extrémité 5' du domaine conservé et ont des extrémités 3' se prolongeant ou non jusqu'au codon stop naturel de LFY (pCH22 et pCH23 à partir du vecteur pIVEX2.4d, pCH26 et pCH27 à partir du vecteur pETM-11) et deux possèdent des paires de bases supplémentaires en amont du domaine conservé et ont des extrémités 3' se prolongeant ou non jusqu'au codon stop naturel de LFY (pCH24 et pCH25 à partir du vecteur pIVEX2.4d, pCH28 et pCH29 à partir du vecteur pETM-11; Fig.18B).

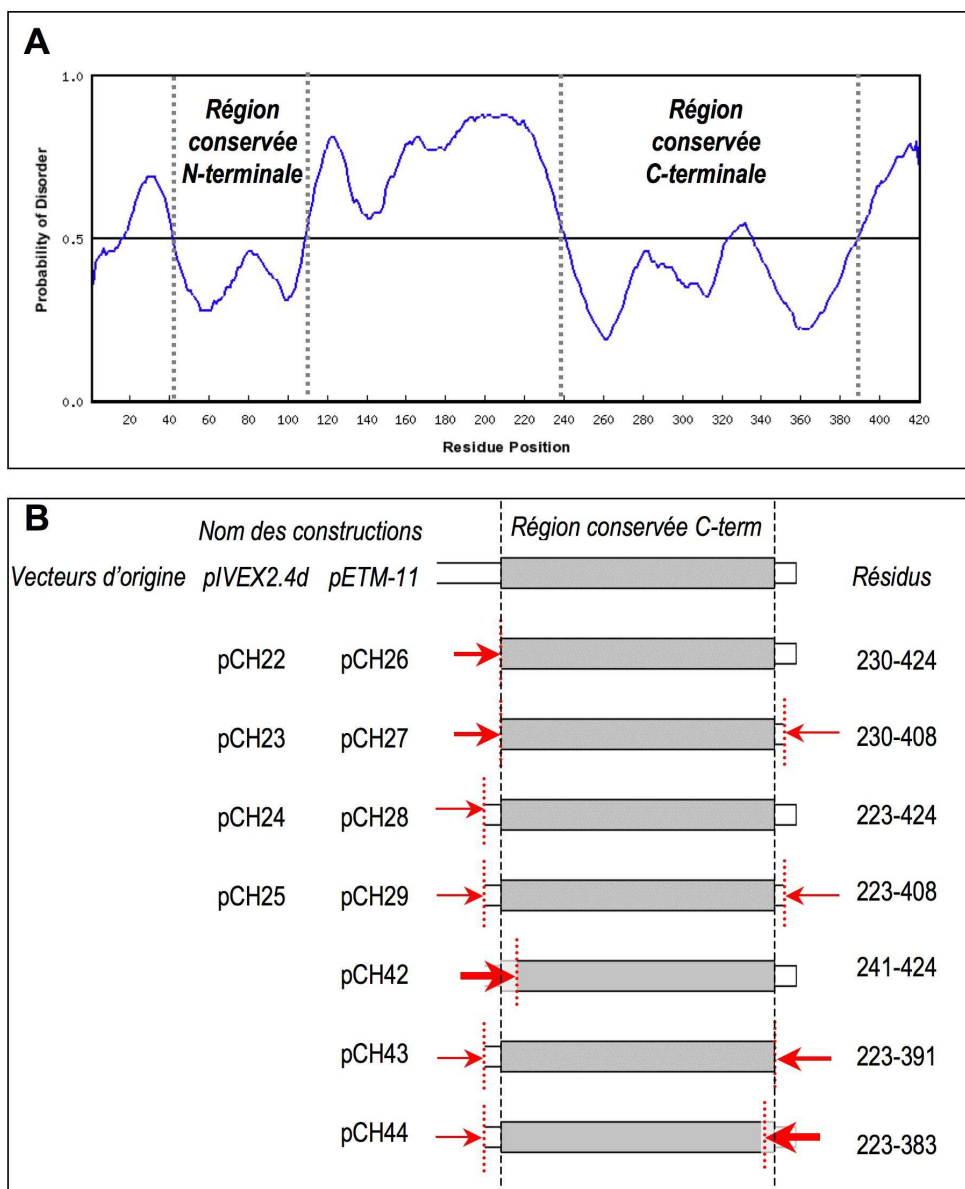


Figure 18: Analyse du domaine C-terminal de LFY

(A) Exemple de prédiction du désordre intrinsèque de la protéine à l'aide du logiciel RONN (Regional Order Neural Network; http://www.strubi.ox.ac.uk/cgi-bin/disorder_results_jan2005.cgi) de la division de biologie structurale de l'Université d'Oxford (STRUBI). Les régions structurées correspondent aux zones où la probabilité de désordre est inférieure 0.5, i.e. les régions 1-17, 43-109, 241-324 et 336-389; (B) Récapitulatif des domaines C-terminaux de LFY clonés. Les flèches rouges soulignent les truncations effectuées.

Le premier essai d'expression a été entrepris par système TNT (transcription traduction *in vitro*, système de transcription rapide RTS, Roche) à partir du vecteur pIVEX2.4d. Aucune expression n'ayant été observée quelle que soit la construction considérée (Fig.19A), je me suis orientée vers l'essai de production chez les bactéries *E. coli* à partir du vecteur pETM-11. Contrairement au premier essai, l'expression de pETM-11 en *E. coli* permet d'observer une très belle induction de l'expression de protéines (de l'ordre de 50mg/L de culture) qui se révèlent en large partie solubles pour les quatre constructions testées (Fig.19B). Une étape de chromatographie d'affinité pour l'étiquette N-terminale 6his (résine *Ni-sepharose High performance*, GE Healthcare) montre en revanche que les produits des

constructions pCH27 et pCH29 se comportent différemment de ceux des constructions pCH26 et pCH28, présentant deux bandes d'importance égale sur gel d'acrylamide SDS-PAGE contre une seule pour ceux de pCH26 et pCH28 (Fig.19C). Ces deux bandes, pouvant signaler des arrêts de traduction ou une plus grande sensibilité des produits d'expression de pCH27 et pCH29 à des protéases, nous ont conduit à ne sélectionner que les produits d'expression des pCH26 et pCH28 pour la suite de nos expériences.

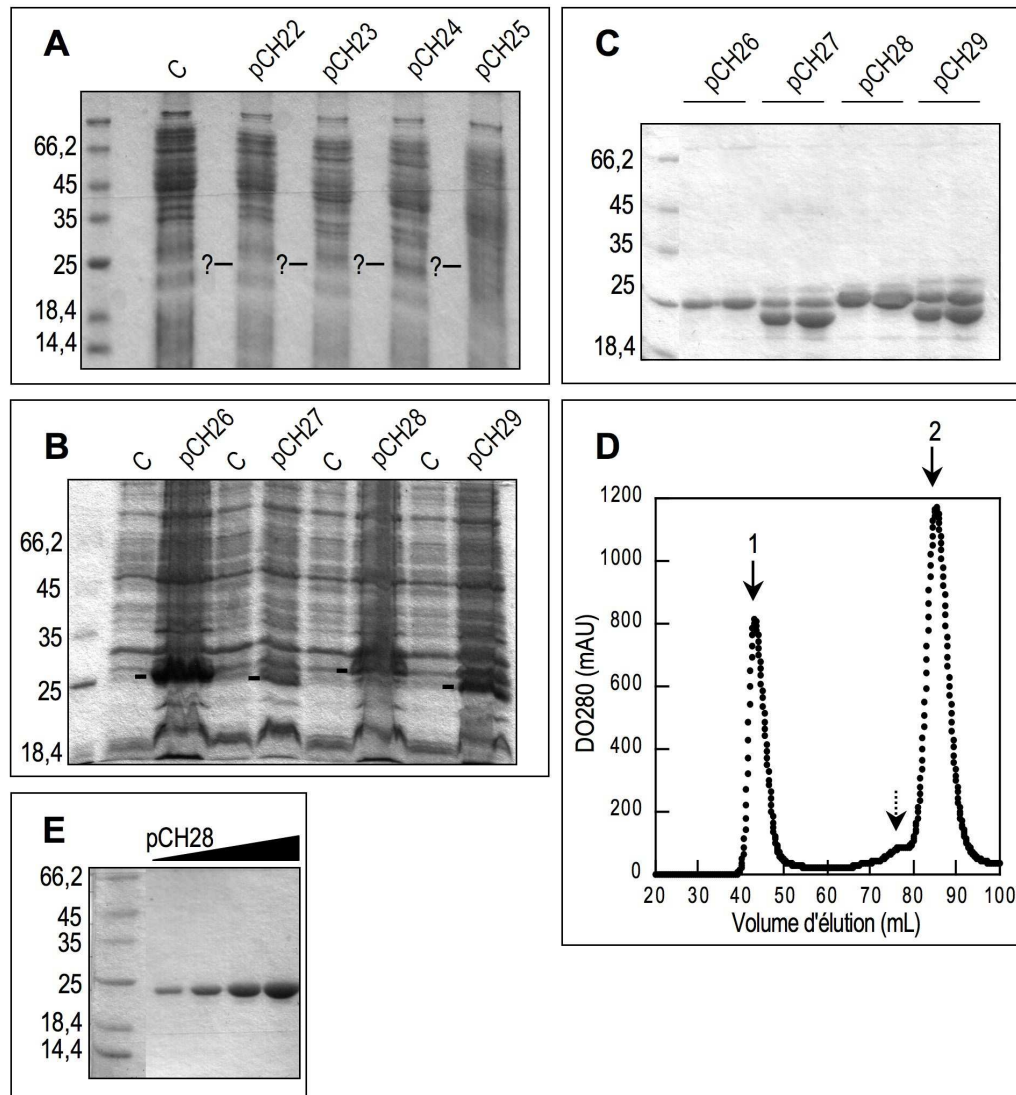


Figure 19: Récapitulatifs des étapes de purification du domaine C-terminal de LFY

(A-B) Evaluation de l'induction par gel SDS-PAGE (A) par système TNT à partir des vecteurs pCH22 à pCH25 et (B) par expression chez *E. Coli* à partir des vecteurs pCH26 à pCH29. La gamme obtenue avec les protéines standard est signalée à gauche (avec masse moléculaire en kDa). La migration attendue ou observée de la protéine d'intérêt est signalée d'un trait noir. Les contrôles 'C' correspondent au dépôt d'une fraction non induite; (C) Analyse par gel SDS-PAGE des fractions d'élution de la chromatographie d'affinité pour les 6his recueillies à partir des expressions de pCH26 à pCH29 (5 et 8 μ L déposés dans chaque cas); (D) Profil de SEC (DO_{280}) des protéines issues du vecteur pCH28 (la flèche 1 signale la fraction agrégée, la flèche 2 la fraction non agrégée, la flèche en pointillés la présence d'un léger pic précédant le pic principal de fraction non agrégée); (E) Analyse sur gel SDS-PAGE d'une quantité croissante (0.5, 0.8, 1, 1.5 μ L) de protéines obtenues à partir de la concentration des fractions non agrégées.

Après une deuxième étape de purification par chromatographie d'exclusion de taille (SEC; colonne Superdex-200, 16/60 GE Healthcare) permettant d'éliminer la proportion de protéines agrégées, les protéines finales issues de pCH26 et pCH28 se révèlent pures et abondantes, résistent au clivage à la protéase TEV (*Tobacco Etch Virus*) de leur étiquette 6his, et supportent une concentration pouvant atteindre plus de 15mg/mL pour un rendement de purification final de l'ordre de 3% (soit environ 1,5mg/L de culture initial; Fig. 19D et E).

Sachant qu'une seule construction de référence nous suffisait pour étudier les propriétés du domaine C-terminal, et que pCH28 permet de produire une protéine présentant 7 aa en plus de pCH26 à l'extrémité N-terminale, nous avons retenu ce vecteur pCH28. Cette protéine produite à partir du vecteur pCH28 (résidus 223-424) sera appelée 'LFY-C' par la suite.

Pour l'étude relative des protéines LFY-C sauvage et mutantes dans le cadre de l'article Hamès et collaborateurs (Annexe 1), les études de liaison à l'ADN ont toutes été réalisées dans les mêmes conditions, à savoir à partir d'une protéine 'fraîche' (venant d'être produite) et dosée par technique de Bradford (Bradford, 1976). L'importance de la fraîcheur de la protéine a été déduite de la comparaison d'affinité de protéines néo-produites, ayant séjournées plus de deux jours à 4°C, ayant été congelées à -20°C en présence de 50% de glycérol, ou encore ayant été congelées à -80°C. Même si dans les trois derniers cas cités, les protéines sont encore fonctionnelles, une diminution claire de leur affinité pour l'ADN comparativement aux protéines néo-produites est détectée. Le choix du dosage Bradford quant à lui est arbitraire. Nous avons pratiqué au laboratoire différentes techniques de dosage protéique (dosage sur gel d'acrylamide dénaturant par rapport à une gamme témoin; dosage par mesure de l'absorption à 280nm et confrontation à l'indice d'absorption spécifique de LFY-C; dosage Bradford) et nos résultats ont révélé que d'une technique à l'autre pour un même échantillon, des écarts d'un facteur d'ordre 2 pouvaient être constatés. Il était donc nécessaire de convenir d'une unique méthode de dosage pour avoir une référence commune.

3. Fonctionnalité et caractérisation quantitative de la liaison de LFY-C à l'ADN

Après avoir produit le domaine LFY-C, il est essentiel de tester s'il est fonctionnel. Le seul critère dont nous disposons est sa capacité à lier l'ADN de ses séquences cibles. Il existe pour cela plusieurs techniques. Nous en avons choisi deux en tirant parti d'un marquage fluorescent qui évitait l'utilisation de la radioactivité. La fonctionnalité de la protéine LFY-C a ainsi été validée par anisotropie de fluorescence (AF) et EMSA, deux techniques complémentaires, à partir d'oligonucléotides marqués au fluorochrome TAMRA contenant un de ses sites de liaison au niveau des séquences régulatrices d'AP1 ou AG (Fig.12, oligonucléotides AP1, AG-I et AG-II). Par souci de clarté, j'entendrai par 'AP1' la notion d'"oligonucléotide AP1".

a) Principe des techniques

1- L'anisotropie de fluorescence (AF)

La technique d'AF (Lundblad et al., 1996) repose sur l'étude de la mobilité rotationnelle d'une molécule marquée fluorescente (un oligonucléotide dans notre cas). Si la molécule marquée est libre, elle aura une grande mobilité rotationnelle, la valeur d'AF sera faible. Au contraire, si cette même molécule est complexée (liée à LFY-C), sa mobilité rotationnelle sera plus réduite et la valeur d'AF plus forte (Fig.20A). L'AF est par définition (voir matériel et méthodes) comprise entre 0 et 1, valeurs limites théoriques pour une molécule à mouvement infiniment rapide (0) ou complètement immobile (1).

2- L'expérience de retard sur gel (EMSA)

La technique d'EMSA repose sur l'étude du profil de migration d'un oligonucléotide (marqué par une sonde fluorescente dans notre cas) sur gel d'acrylamide en conditions non dénaturantes. Si l'oligonucléotide est libre, il migrera rapidement. En revanche, s'il est lié à une protéine, il sera 'retardé'.

3- La résonance plasmonique de surface (SPR)

La technique de SPR utilise les propriétés optiques d'une interface transparente (du verre par exemple) entre deux milieux d'indices de réfraction différents. Selon l'angle d'incidence d'un faisceau de lumière polarisée monochromatique illuminant cette interface, soit une partie de la lumière incidente est réfléchi et l'autre est réfractée, soit toute la lumière incidente est réfléchi, et seule une composante électromagnétique de la lumière (l'onde évanescente) se propage perpendiculairement à l'interface sur une distance équivalente à sa longueur d'onde. Lorsqu'une couche fine d'or riche en électrons libres est rajoutée sous le verre et est traversée par cette onde évanescente, les électrons du métal entrent en résonance. Ce phénomène est appelé la 'résonance plasmonique de surface'. Si le faisceau réfléchi est analysé à l'aide d'un détecteur optique (une barrette de diodes), une chute d'intensité à un angle défini est mise en évidence. Cet angle d'intensité minimum se nomme l'angle de résonance et il a la propriété de varier en fonction de l'indice de réfraction du milieu (Fig.20B).

La technique de SPR permet de visualiser des interactions en temps réel et sous un débit continu de tampon entre des biocapteurs (ADN de choix greffé à une couche de dextran accolée à la couche d'or) et des analytes (protéine LFY-C dans notre cas). Les changements de masse induits par l'association ou la dissociation des complexes protéines/ADN modifient l'indice de réfraction du milieu et décalent la position de l'angle de résonance. L'enregistrement de la variation de l'angle de résonance permet de suivre en temps réel la fixation des molécules injectées sur le biocapteur.

b) Test de fonctionnalité et estimation de la constante de dissociation de LFY-C pour AP1

Les tests d'interaction par AF et EMSA entre LFY-C et l'ADN cible ont été réalisés à partir d'oligonucléotides de 30pb.

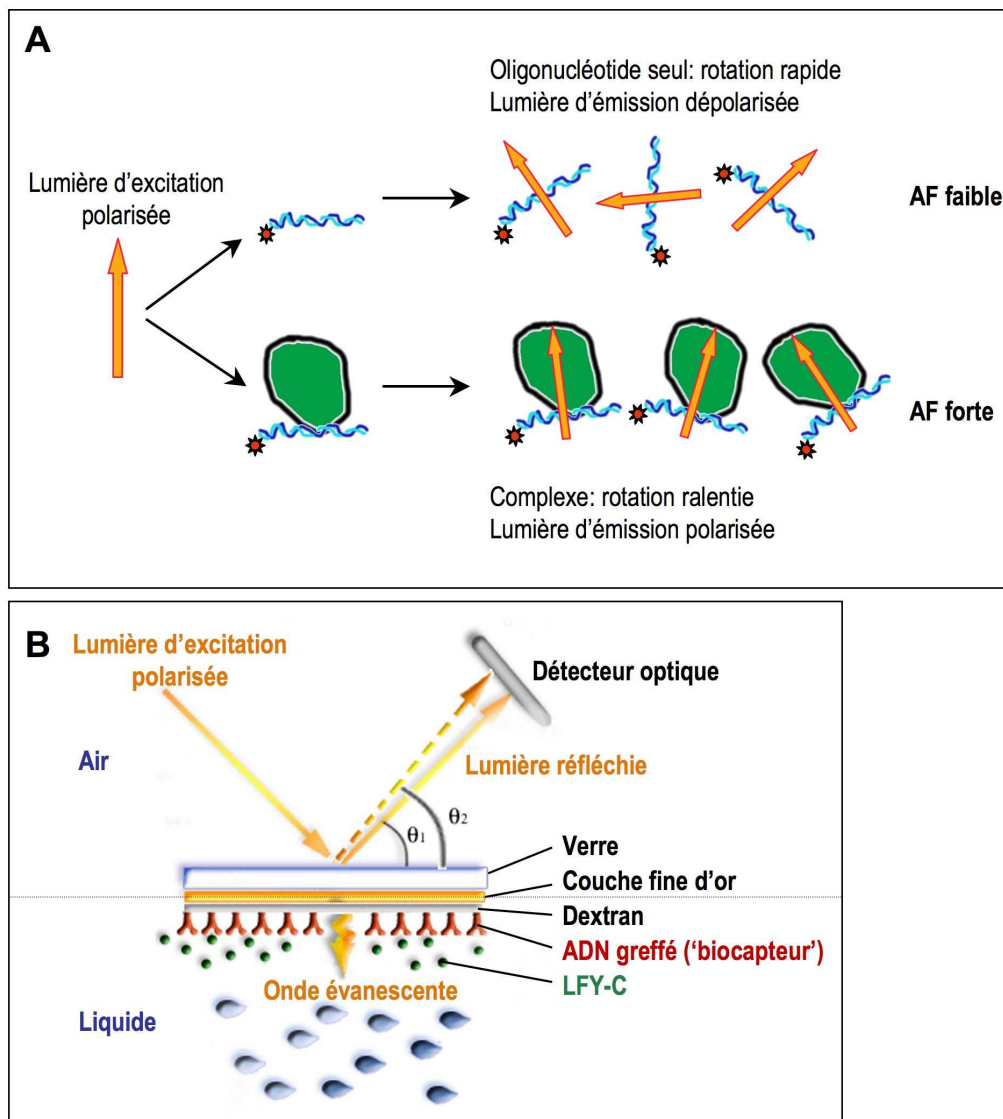


Figure 20: Description des techniques d'anisotropie de fluorescence (AF) et de résonance plasmonique de surface (SPR)

(A) Représentation schématique de la technique d'AF. La valeur d'AF est proportionnelle à : $[(\text{intensité de l'émission verticale}) - (\text{intensité de l'émission horizontale})] / (\text{intensité totale émise})$ pour une excitation par une lumière polarisée verticalement; (B) Représentation schématique de la technique de SPR. La position de l'angle de résonance θ_2 varie en fonction de l'association LFY-C/ADN.

La figure 21 (A) montre les valeurs d'AF de 10nM d'AP1 en fonction d'une concentration croissante de protéines LFY-C. On constate que les valeurs d'AF augmentent avec la concentration de LFY-C pour atteindre un plateau indiquant une saturation des oligonucléotides par LFY-C. Notre protéine est donc capable de lier l'ADN.

La confrontation de l'équation théorique (décrite en Matériel et Méthodes) aux résultats expérimentaux à l'aide du logiciel KaleidaGraph nous permet d'estimer la constante de dissociation (K_d) de la réaction à environ 100nM.

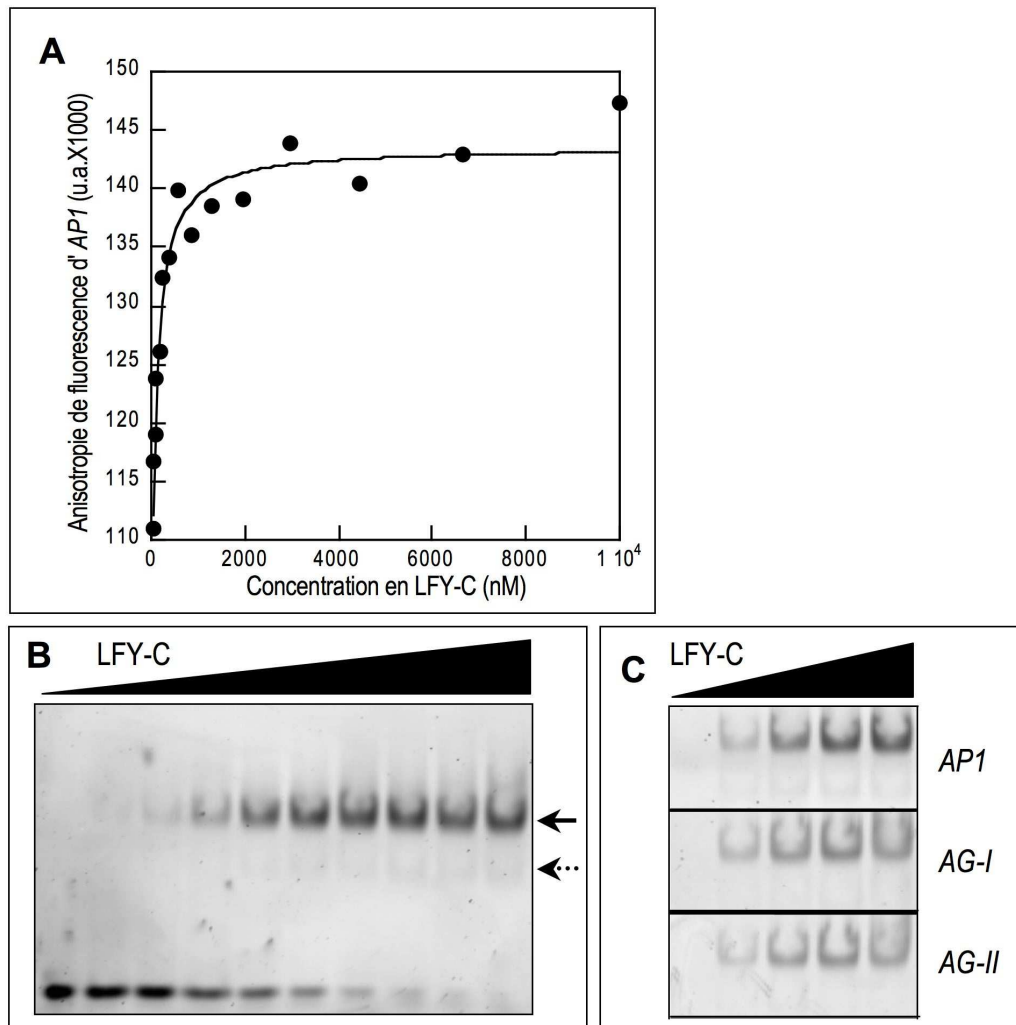


Figure 21: Test de fonctionnalité de LFY-C

(A) Courbe de saturation de liaison de LFY-C à AP1 WT obtenue par AF; (B) EMSA réalisé avec 10nM AP1 et une concentration croissante de LFY-C (0, 10, 20, 50, 100, 200, 500, 1000, 2000, 3000nM de gauche à droite). La flèche continue signale le complexe majoritaire, la flèche en pointillés le complexe minoritaire; (C) EMSA réalisé avec 10nM d'oligonucléotides (nature précisée) et une gamme croissante de LFY-C (0, 100, 250, 750, 2000).

Suite au mélange d'AP1 et de LFY-C, l'expérience d'EMSA révèle l'apparition de deux nouveau complexes retardés qui se distinguent de l'ADN seul. Nous ne retiendrons ici que la bande retardée majoritaire qui présente une intensité augmentant avec [LFY-C] jusqu'à atteindre un plateau (flèche continue noire, fig.21B).

Après avoir quantifié les bandes correspondantes à l'ADN libre et à l'ADN retardé en fonction de [LFY-C], le K_d de LFY-C pour AP1 est estimé par cette technique à environ 100nM également (détaillé en matériel et méthodes). Mentionnons que l'affinité de LFY-C pour deux autres séquences cibles, AG-I et AG-II, semble très similaire (Fig.21C).

c) Estimation de la constante d'affinité de LFY pour des régions régulatrices entières par SPR (travail d'E.Moyroud)

Edwige Moyroud, étudiante en thèse au sein de notre équipe, a mis en place en collaboration avec l'équipe de Charlie Scutt (RDP, ENS-Lyon) une technique de SPR permettant de tester la liaison de LFY à des grands fragments d'ADN (jusqu'à 3kb). Ses résultats sont reportés en Annexe 2 (article soumis). Elle a pour cela testé la capacité de LFY à interagir avec des séquences cis régulatrices d'*AP1* (2,2kb), *AP3* (2,9kb) et *AG* (3kb), gènes cibles connus de LFY, et avec celles de *HSP70* (2,4kb) et *BHLF1* (0,7kb) à titre de contrôle négatif. Par cette approche, LFY lie la région cis régulatrice d'*AP1* avec une affinité apparente de $15 \pm 0,9\text{nM}$, celle d'*AP3* avec une constante de dissociation apparente (K_d) de $21,4 \pm 2,6\text{nM}$, celle d'*AG* avec une affinité apparente de $0,6 \pm 0,02\text{nM}$. Ces résultats se révèlent extrêmement discriminants par rapport à ceux obtenus avec les contrôles négatifs, sachant que LFY lie les régions cis régulatrices de *HSP70* et *BHLF1* avec une affinité apparente de plus de 1000nM.

Ces valeurs de K_d de l'ordre de 1 à 100nM sont cohérentes avec les K_d déjà décrits dans la littérature pour d'autres facteurs de transcription, de valeur comprises entre 1nM et 700nM, signalant la validité biologique de ces résultats (Hurst, 1995).

d) Influence de la longueur de l'oligonucléotide sur l'affinité de liaison de LFY-C à l'ADN

Pour les deux expériences précédentes, nous avons utilisé des oligonucléotides de 30pb. Or la séquence consensus décrite dans la littérature ne contient que 7pb. Pour savoir si les paires de bases contactées par LFY s'étalent sur ces 30pb ou si elles se limitent à un domaine plus restreint comme pourrait nous le suggérer sa séquence consensus, nous avons étudié l'influence de la longueur de l'oligonucléotide sur l'affinité de liaison de LFY à l'ADN. Nous avons pour cela réalisé des expériences de compétition en AF entre l'oligonucléotide *AP1* de 30pb fluorescent et des oligonucléotides *AP1* de 30pb (oligonucléotide de référence), de 25pb ou de 21pb non marqués (Fig.22). Sans quantifier cette perte d'affinité, nous constatons que les compétitions des oligonucléotides non marqués de 25pb et 30pb avec les oligonucléotides de 30pb fluorescents sont pratiquement équivalentes, tandis que les oligonucléotides de 21pb entrent nettement moins bien en compétition (perte plus progressive d'AF): LFY-C a une meilleure affinité pour les oligonucléotides de 25pb et 30pb que pour les oligonucléotides de 21pb.

Ces résultats révèlent donc que LFY-C contacte sans doute l'ADN sur au moins 22pb.

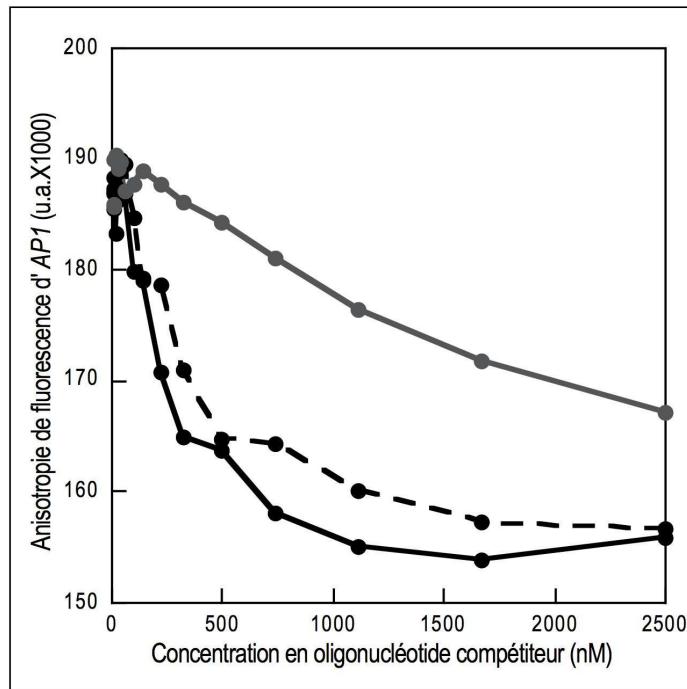


Figure 22: Etude par compétition en AF de l'affinité de LFY-C pour des oligonucléotides de 21, 25 et 30pb

La concentration initiale en LFY-C pour cette expérience est de 500nM, celle des oligonucléotides marqués fluorescents (AP1 30pb) de 10nM. Courbes de compétition avec les oligonucléotides de 30pb en noir, avec les oligonucléotides de 25pb en pointillés noirs, et avec les oligonucléotides de 21pb en gris.

4. Domaine minimal de liaison de LFY à l'ADN

Ayant confirmé que notre protéine produite était bien fonctionnelle, nous avons délimité précisément le domaine minimal de liaison LFY à l'ADN décrit par Maizel et collaborateurs (2005). Dans leur article, ces auteurs le délimitent aux résidus 230 à 424.

Nous avons dans un premier temps confirmé la délimitation en N-terminal en comparant la liaison d'une protéine débutant au ras du domaine conservé (résidus 230-424, vecteur pCH26; Fig.18B) avec une protéine tronquée des 11 premiers acides aminés en position N-terminale (résidus 241-424, vecteur pCH42; Fig.18B). Si les protéines produites à partir de pCH26 lient normalement l'ADN par EMSA, celles issues de l'expression de pCH42 présentent une affinité pour l'ADN fortement diminuée (Fig.23). Nos résultats sont donc en accord avec les résultats de Maizel et collaborateurs délimitant l'extrémité N-terminale du domaine minimal de liaison de LFY à l'ADN à l'extrémité du domaine conservé C-terminal (résidu 230).

Dans un second temps, nous avons étudié si la région C-terminale présentant un caractère légèrement hydrophobe (Fig.18A région C-terminale de probabilité de désordre supérieure à 0,5; résidus 392-424) était importante à l'interaction LFY/ADN. Pour cela, j'ai testé deux autres constructions: une protéine tronquée au ras de l'extrémité conservée C-terminale (résidus 223-391, vecteur pCH43; Fig.18B), et une protéine tronquée des 9 résidus C-terminaux de LFY (résidus 223-383,

vecteur pCH44; Fig.18B). Par EMSA (Fig.23), nous constatons que les protéines produites à partir de pCH43 sont toujours capables de lier l'ADN, alors que celles issues de pCH44 n'en sont plus capables.

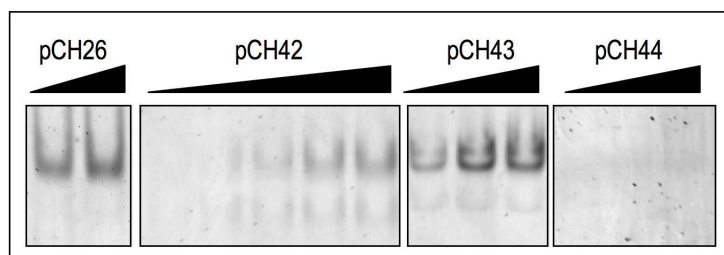


Figure 23: Détermination du domaine minimal de liaison de LFY-C à l'ADN par EMSA

Tous les EMSA sont réalisés avec 10nM *AP1* et une concentration variable en protéines obtenues à partir de l'expression des vecteurs pCH26 (275, 550nM de protéines), pCH42 (1000, 2000, 3000, 4000, 5000nM de protéines), pCH43 (100, 500, 1000nM de protéines) et pCH44 (1000, 2000, 5000nM de protéines).

Ces résultats indiquent donc que les résidus 383 à 391 jouent un rôle important dans la capacité de liaison de LFY à l'ADN, mais qu'en revanche, la région hydrophobe C-terminale n'est pas nécessaire. **Le domaine minimal de liaison de LFY à l'ADN serait donc délimité par les résidus 230 à 391.** Notons que dans leur article, Parcy et collaborateurs (1998) étudient par EMSA les propriétés de liaison d'une protéine délétée de la région hydrophobe (résidus 391-424, Fig.13A), truncation qui, d'après ces auteurs, améliore l'expression protéique en bactéries. Leurs données, montrant que cette protéine recombinante reconnaît spécifiquement *AP1*, suggère que cette délimitation C-terminale du domaine minimal de liaison soit vraie également dans le contexte de la protéine LFY entière (notée LFY FL, *Full Length*).

Il aurait été intéressant de valider ces résultats par une analyse de repliement de la protéine, en dichroïsme circulaire ou en suivi de la fluorescence tryptophane par exemple. Il serait en effet possible que ces protéines adoptent un mauvais repliement qui leur fasse perdre leur fonctionnalité. Rappelons que les résidus 230 à 391 correspondent aux limites du domaine conservé C-terminal, domaine prédit comme structuré. Par conséquent, la truncation de résidus à ses extrémités pourrait le déstabiliser.

5. Détermination de la nature des deux complexes de LFY-C/ADN

En regardant plus attentivement l'EMSA en figure 21(B), nous avons constaté que bien que le complexe retardé décrit initialement soit très majoritaire, un autre complexe de plus mobilité est également visible: LFY-C forme donc deux complexes de ratio protéine(s):ADN différents dont j'ai voulu déterminer la nature.

L'identification des deux complexes LFY/ADN exige de déterminer la masse moléculaire de chacun de ces complexes. Si d'apparence cette détermination pourrait sembler facilement accessible, nous allons voir que dans notre cas, contraints de travailler en conditions non dénaturantes, nous nous sommes heurtés à quelques difficultés. Plusieurs expériences biochimiques ont été menées à ces fins.

a) Ferguson Plot

Le principe de la technique de Ferguson Plot est d'étudier le comportement de migration de molécules ou complexes sur une gamme de gels natifs à différentes concentrations en acrylamide (de 5 à 11%; fig.24A-C). Les mesures de la distance de migration du complexe principal LFY-C/AP1 dans le cas présent sont réalisées à partir de chaque gel et confrontées aux mesures d'une gamme de protéines standards dont les masses moléculaires sont connues (détaillée en matériel et méthodes).

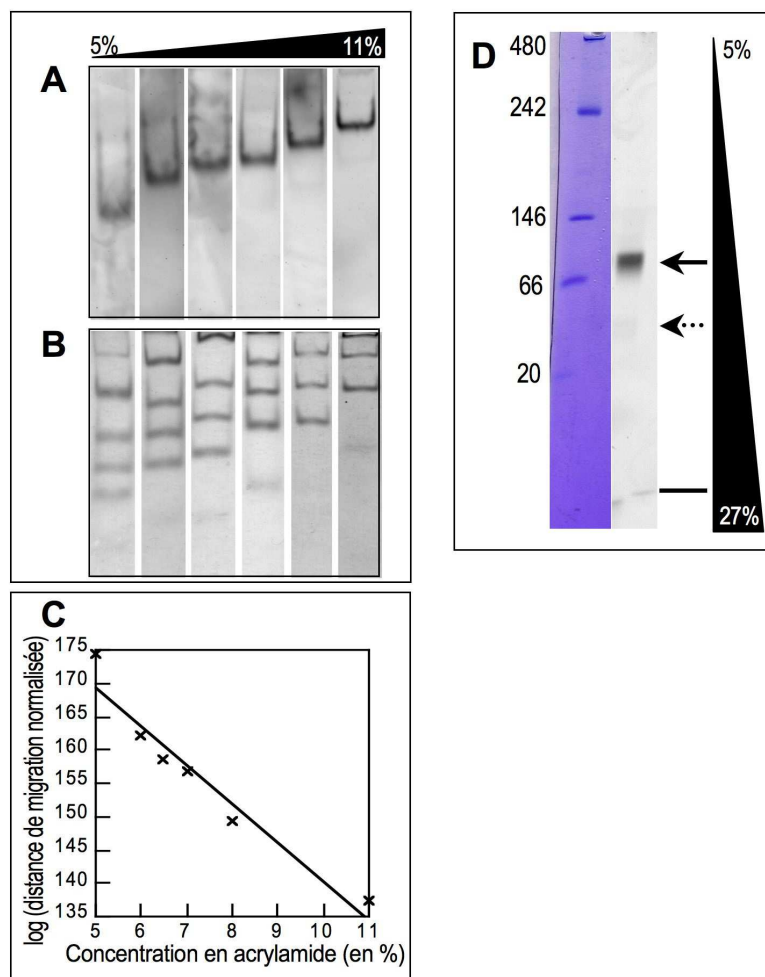


Figure 24: Etude de la nature des deux complexes LFY-C/ADN par gel natif

(A) à (C) **Par Ferguson plot.** (A) Expérience réalisée avec 10nM AP1 et 1500nM LFY-C. Bilan des profils de migration du complexe majoritaire sur gel natif 5, 6, 6.5, 7, 8 et 11% acrylamide; (B) Bilan profils de migration des protéines de référence sur gel natif 5, 6, 7, 8, 9, 11% acrylamide; (C) Représentation graphique de $\log(\text{distance de migration du complexe majoritaire rapportée au front de migration}) = f(\% \text{ d'acrylamide})$ permettant d'estimer une pente; cette valeur permet ensuite d'estimer la masse du complexe à partir de la droite étalon établie avec les protéines standards (Matériel et Méthodes); (D) **Par EMSA gradient à l'équilibre (GEMSA).** Expérience réalisée avec 10nM AP1 et 1800nM LFY-C. La migration de l'oligonucléotide fluorescent (gel de droite) est visualisée à l'aide du fluorimager Typhon 9400 (Molecular dynamics) la migration des protéines standard (gel de gauche, masse moléculaire précisée en kDa) est visualisée par coloration au bleu de coomassie. La flèche solide indique le complexe principal, la flèche en pointillés indique le complexe minoritaire, et le trait indique l'oligonucléotide seul. Le gradient d'acrylamide est schématisé à droite.

Cette technique a conduit à une estimation de la masse moléculaire du complexe principal LFY-C/AP1 en moyenne à 80kDa sur 4 expériences indépendantes (j'ai volontairement exclue une valeur aberrante à 140kDa). Sachant que LFY-C (produit à partir de pCH28) a une masse moléculaire de 25,7kDa et

AP1 une masse moléculaire de 13kDa, le Ferguson plot suggère donc que le complexe principal comprenne donc entre 2 et 3 protéines LFY-C par ADN.

b) EMSA sur gel gradient à l'équilibre (GEMSA)

Le gel gradient à l'équilibre permet d'étudier la masse moléculaire de protéines ou complexes protéiques en conditions non dénaturantes. Nous l'avons adapté pour l'étude de la masse moléculaire d'un complexe protéine/ADN par EMSA. Il consiste à faire migrer les espèces en présence pendant plusieurs heures sur un gel natif réalisé à partir d'un gradient de concentrations en acrylamide. Quand elles atteignent un front où l'acrylamide est trop concentré pour leur permettre de continuer leur migration, elles s'immobilisent. Cette concentration limite en acrylamide est fonction de la masse moléculaire et de la conformation de la molécule ou du complexe considéré(e). La distance de migration du complexe LFY-C/*AP1* est ensuite comparée à celles de protéines de référence (Fig.24D et Matériel et Méthodes). Du fait d'une variabilité inexplicée dans la migration du complexe, cette technique a estimé à 3 reprises que le complexe principal se composait de 2 protéines LFY-C par ADN, et à une reprise qu'il se composait de 4 LFY-C par ADN.

c) Chromatographie par exclusion de taille (SEC) analytique

La technique de SEC permet de séparer des molécules ou des complexes en fonction de leur volume hydrodynamique (qui dépend de leur masse et leur conformation en solution). Elle repose sur les propriétés d'une résine constituée de billes à pores d'une taille telle qu'ils retiennent plus longtemps les molécules les plus petites. Les molécules les plus volumineuses sont donc éluées plus tôt. Afin de déterminer la masse du complexe LFY-C/ADN, nous avons utilisé une résine plus résolutive (Superdex-200 10/300 GL; GE Healthcare) que celle utilisée pour la purification de la protéine. Après avoir calibré la colonne à partir de protéines de référence, nous avons successivement analysé LFY-C seule, *AP1* seul, et le mélange LFY-C/*AP1* (Fig.25A et Matériel et Méthodes). L'analyse de la protéine seule, éluant en un pic à 28 kDa, est en accord avec la masse moléculaire d'une protéine LFY-C (25,7kDa) et révèle donc **qu'en solution, LFY-C est majoritairement monomérique**. *AP1* seul quant à lui élue à une masse moléculaire aberrante de 52kDa (masse théorique 13kDa), signalant un biais important de l'analyse de cette molécule par SEC, conséquence très probable de sa forme allongée. L'analyse du mélange LFY-C/*AP1* révèle un pic à 121 kDa. Connaissant l'importance de la conformation tridimensionnelle dans l'analyse SEC, il nous est difficile de tirer une conclusion de ce résultat. Si l'on considère qu'*AP1* au sein du complexe se comporte comme une molécule de masse moléculaire égale 13kDa, le complexe contient $[(121-13)/28] \approx 3.9$ LFY-C par *AP1*, alors que s'il se comporte comme une molécule de 52kDa, le complexe contient $[(121-52)/28] \approx 2.5$ LFY-C par *AP1*.

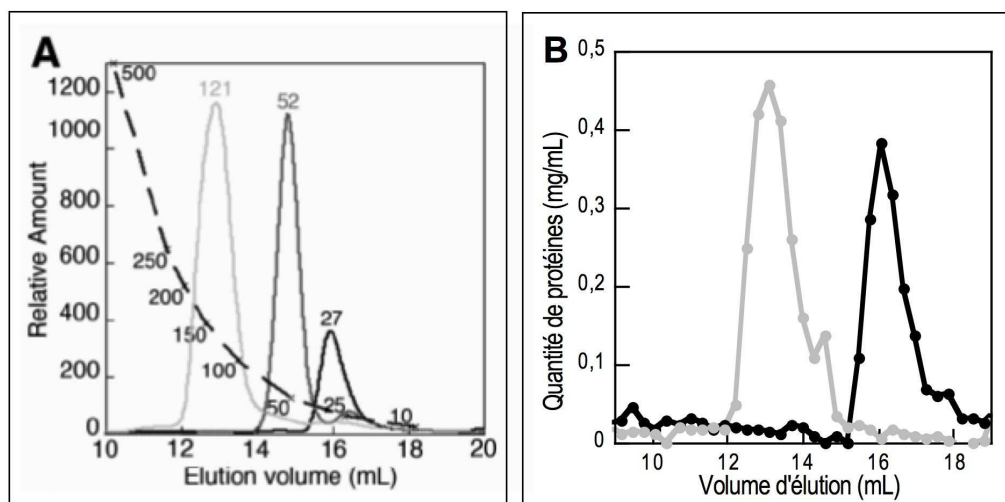


Figure 25: Etude de la nature des deux complexes LFY-C/ADN par SEC analytique

(A) Profil de SEC (DO_{280}) de LFY-C (40 μM, courbe noire), AP1 DNA (10 μM, courbe gris foncé), LFY-C (40 μM) + AP1 DNA (10 μM, courbe gris claire). Les poids moléculaires estimés à partir de la courbe de calibration (courbe en pointillés noirs) sont indiqués; (B) Quantification de LFY-C seule (courbe noire) ou analysée avec AP1 (courbe gris claire) par dosage Bradford à partir des fractions d'élution de SEC.

d) Premier bilan: LFY se lierait à l'ADN selon un ratio 4LFY-C/ADN ?

Cette première série d'expériences nous indique qu'il y a entre 2 à 4 LFY-C par ADN, mais aucun argument ne permet d'établir un ratio précis.

En parallèle de ces expériences, nous avons également entrepris des expériences étudiant la liaison de LFY-C à AP1 d'un point de vue quantitatif. Dans le cas de l'expérience de SEC décrite ici par exemple, l'expérience a été réalisée à partir de 40μM de LFY-C et 10μM d'AP1. Le fait que la protéine soit entièrement déplacée par ajout d'un ratio ¼ d'ADN (résultat confirmé par dosage Bradford, fig.25B) apportait un argument fort en faveur de l'hypothèse 4 LFY-C par ADN. Nous allons voir qu'il est pourtant faux, problème corrélé à la difficulté du dosage de la protéine et de l'ADN auquel je reviendrai dans la discussion de ce chapitre.

L'expérience de SEC analytique est la seule de cette première série d'expériences à avoir été retenue pour la publication de nos résultats car elle apporte une donnée importante, à savoir que LFY-C en absence d'ADN est monomérique. La série d'expériences qui suit est quant à elle entièrement intégrée à l'article Hamès et collaborateurs (Annexe 1) à l'exception des résultats d'ultracentrifugation analytique.

e) Ultracentrifugation analytique (AUC; en collaboration avec Christine EBEL, IBS, Grenoble)

L'AUC allie séparation des macromolécules et étude de la vitesse de sédimentation des espèces présentes en solution et en temps réel. Un important avantage de cette technique est qu'elle ne met en jeu aucune matrice ou autre surface pouvant interagir avec les molécules analysées: le biais

observé dans le cas de l'analyse d'ADN par SEC est complètement supprimé ici. Cependant, dans nos conditions expérimentales, l'analyse s'est révélée délicate. Les données obtenues privilégient l'hypothèse de deux protéines par ADN, mais notre collaboratrice Christine Ebel n'a pas pu l'affirmer avec certitude avec la gamme de concentration de protéines choisie. Par ailleurs, les mesures de DO suggèrent une surestimation aussi bien des quantités de protéines que d'ADN, ramenant la discussion à la difficulté de quantifier précisément ces molécules.

f) MALLS-SEC (en collaboration avec Francine Gérard et Marc Jamin, IVMS, Grenoble)

Le principe du MALLS-SEC est de coupler une SEC classique à un dosage des fractions d'élution à l'aide d'un réfractomètre plus une analyse de diffusion statique de la lumière selon un angle de 360° (MALLS, *Multi-Angle Laser Light Scattering*, Fig.26). Ainsi, si cette technique sépare dans un premier temps les molécules en fonction de leur volume hydrodynamique, elle permet dans un second temps de concilier une analyse de la concentration et des propriétés de diffusion de ces molécules. Les données sont ensuite soumises à une analyse mathématique qui permet de déduire la masse moléculaire du complexe principal LFY-C/AP1 à 64 ± 2 kDa (pour une explication détaillée de l'analyse mathématique, se référer à l'article de nos collaborateurs; Gerard et al., 2007). Ces résultats sont donc en accord avec un ratio 2LFY-C/AP1 dont la masse moléculaire théorique est de 65kDa. Mentionnons néanmoins que l'analyse de la protéine seule propose une masse moléculaire de LFY-C de 35 ± 1 kDa, résultat ne correspondant ni au monomère (25,7kDa), ni au dimère (51,4kDa). D'après F. Gérard, cette incertitude serait la conséquence d'un calibrage moins précis de l'appareil à ces faibles poids moléculaires.

g) EMSA en présence d'un mélange équimolaire de LFY-C de masses différentes

Une dernière technique visant à confirmer les résultats préliminaires d'ultracentrifugation analytique et les résultats de MALLS-SEC a consisté à étudier le comportement de liaison à AP1 de deux protéines LFY-C de masses différentes, l'une fusionnée à la GFP (*Green Fluorescent Protein*, 28kDa) et à une étiquette 6his, l'autre sans autre fusion traductionnelle que l'étiquette 6his. En figure 27, nous pouvons voir que LFY-C/AP1 et LFY-C:GFP/AP1 forment un complexe majoritaire migrant différemment. Suite au mélange équimolaire de LFY-C et LFY-C:GFP avec AP1, une seule bande intermédiaire à ces complexes majoritaires apparaît, révélant la formation d'un unique complexe hybride LFY-C/LFY-C:GFP/ADN. Ce résultat confirme donc l'hypothèse d'un ratio, 2LFY-C par ADN. Les bandes mineures de bas poids moléculaire sont donc interprétées comme des complexes 1LFY-C par ADN.

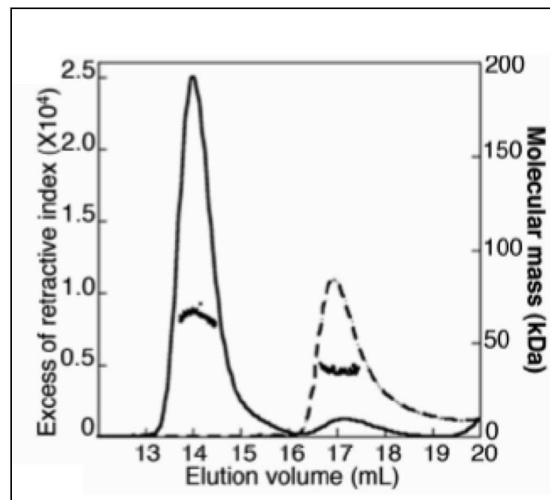


Figure 26: Analyse par SEC-MALLS du complexe LFY-C/AP1

Profil de SEC (excès d'index réfractif, axe des ordonnées gauche) de LFY-C seule (courbe en pointillés noirs) ou avec AP1 (courbe noire). Les points signalent la distribution des masses moléculaires estimées par cette technique (axe des ordonnées droit).

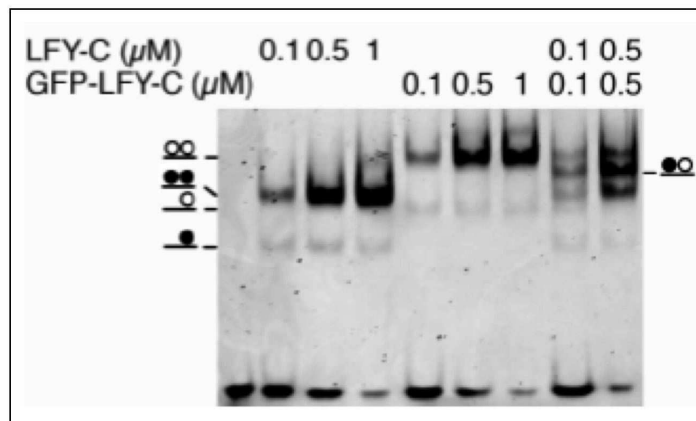


Figure 27: EMSA en présence d'un mélange équimolaire de LFY-C et LFY-C:GFP

L'expérience est réalisée avec 10nM AP1 et des concentrations variables de LFY-C et/ou LFY-C:GFP. Les complexes LFY-C/AP1 et LFY-C:GFP/AP1 sont schématisés respectivement par des cercles noirs et blancs.

h) Deuxième bilan: LFY se lie à l'ADN selon un ratio 2 LFY-C/ADN

Ainsi, cette dernière série d'expériences fournit des arguments tels que nous pouvons admettre que le complexe majoritaire de liaison de LFY-C à l'ADN est constitué de 2 LFY-C par oligonucléotide AP1. Par déduction, le complexe minoritaire correspond à 1 LFY-C par oligonucléotide AP1. Sachant que LFY-C est majoritairement monomérique en solution, le fait que le complexe monomérique existe mais soit toujours minoritaire indique une coopérativité de liaison, point que nous allons détailler par la suite.

6. Mise en évidence d'une coopérativité de liaison de LFY-C à l'ADN

a) Définition de la coopérativité

La coopérativité est un phénomène observé lorsque l'affinité de liaison du premier monomère à l'ADN (complexe C1) est moins bonne que celle du deuxième monomère (complexe C2; Fig.28A). De ce fait, le complexe C2 est toujours majoritaire devant le complexe C1, même à faible concentration en LFY-C.

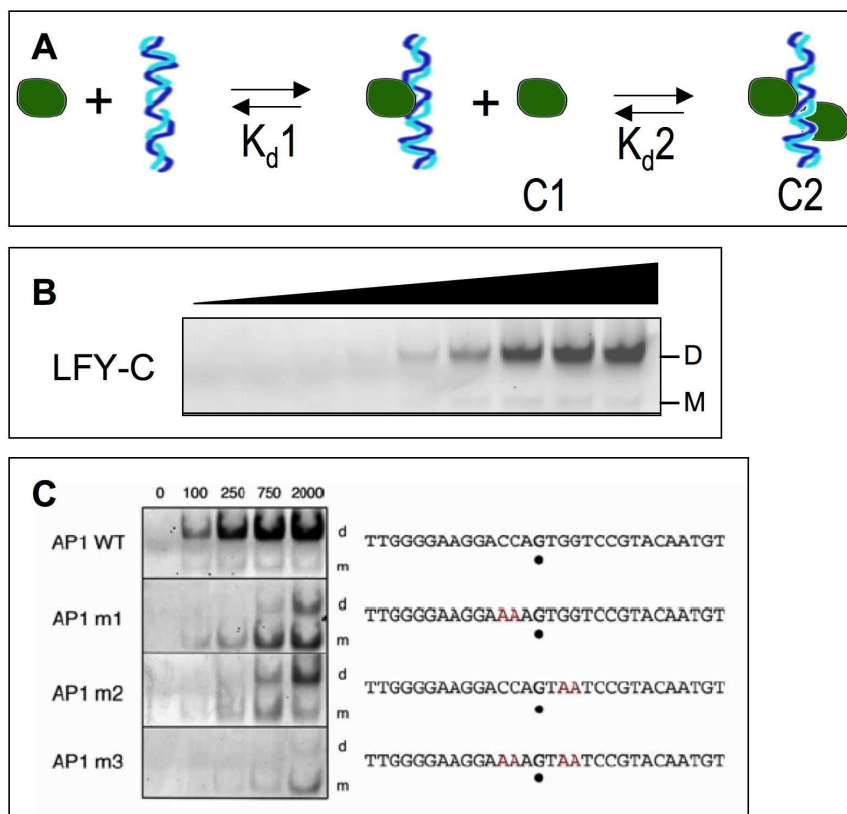


Figure 28: Etude de la coopérativité de liaison de LFY-C à l'ADN

(A) Représentation schématique de l'équilibre réactionnel de liaison de LFY-C à l'ADN. La coopérativité se définit par une valeur de K_d1 supérieure à celle de K_d2 , i.e. une meilleure affinité de liaison du deuxième monomère à l'ADN que du premier monomère; (B) EMSA réalisé avec 50nM AP1 et une concentration croissante en LFY-C (0, 10, 20, 50, 100, 200, 500, 1000, 2000, 3000nM de gauche à droite ; les positions du dimère (D) et du monomère (M) sont indiquées) ; (C) Comparaison d'EMSA réalisés avec 10nM d'oligonucléotides (nature précisée sur la figure, mutations rapportées en rouge; le point noir marque la base médiane).

La coopérativité peut résulter de deux mécanismes moléculaires distincts: soit la liaison de premier monomère augmente l'affinité à l'ADN du second monomère par modification de la conformation de l'ADN; soit le second monomère contacte physiquement le premier monomère et présente donc une affinité de liaison résultant à la fois de son affinité pour le premier monomère et pour l'ADN (Gromiha et al., 2005; Panne et al., 2004; Schumacher et al., 2002; Senear et al., 1998).

La coopérativité est un mécanisme particulièrement intéressant à étudier car il peut intervenir dans la mise en place de transitions développementales brutales, en particulier s'il est conjugué à une boucle d'autorégulation (Burz et al., 1998; Cherry and Adler, 2000). Dans le cas de la transition florale,

LFY joue un rôle déclencheur essentiel alors que son expression augmente graduellement. Sachant que LFY et AP1 s'autorégulent mutuellement (Liljégren et al., 1999; Parcy et al., 1998), y ajouter une coopérativité de liaison de LFY à AP1 pourrait contribuer à expliquer comment LFY agit dans cette transition abrupte.

b) Evaluation de la coopérativité de LFY-C

La coopérativité est évaluée en EMSA à partir d'une gamme croissante de LFY-C à concentration fixe en AP1 (fig.28B). La quantification des différents complexes (dimérique C2, monomérique C1 et ADN libre) permet par un ajustement des courbes au moyen des équations théoriques de déterminer les valeurs de K_{d1} et K_{d2} . On parle de coopérativité lorsque le ratio K_{d1}/K_{d2} est supérieur à 1. De cette manière, l'indice de coopérativité de LFY-C est estimé à 91, traduisant une très forte coopérativité.

c) Influence de la séquence d'ADN sur la coopérativité

Si notre modèle est juste et que LFY se lie effectivement à l'ADN sous forme de deux monomères successifs, on doit pouvoir modifier cette coopérativité de liaison en modifiant la séquence de l'ADN reconnu. Partant du constat que les deux principales cibles de LFY (AP1 et AG) présentent toutes deux des sites de reconnaissance par LFY en partie palindromique (Fig.12), nous avons étudié le comportement de liaison de LFY-C à trois oligonucléotides AP1 mutés, AP1m1 étant muté au niveau de 2pb dans le demi-palindrome 5', AP1m2 au niveau de 2pb dans le demi-palindrome 3', et AP1m3 à la fois au niveau de 2pb dans les demi-palindrome 3' et 5' (Fig.28C). Il n'a pas été nécessaire de calculer l'indice de coopérativité dans ces trois cas de figure pour conclure que la coopérativité était altérée: aussi bien avec AP1m1, AP1m2 et AP1m3, nous constatons qu'aux plus faibles concentrations, le complexe C1 devient majoritaire. La coopérativité est donc étroitement dépendante de la séquence en ADN, et nos résultats laissent supposer que chaque monomère pourrait contacter un demi-palindrome. Relevons que LFY-C n'aurait néanmoins pas exactement le même comportement sur AP1m1 qu'AP1m2: le premier monomère s'accumule plus fortement avec l'oligonucléotide AP1m1 qu'avec AP1m2 alors que le même type de mutation a été introduit dans les deux demi-sites.

En conclusion, nos données sont cohérentes avec l'hypothèse selon laquelle LFY-C lie l'ADN sous forme de dimère avec une liaison coopérative de deux monomères successifs.

7. Etude de la coordination d'ion(s) métallique(s)

a) Les indices à l'origine de cette hypothèse

En analysant l'alignement des séquences des différents orthologues de LFY, plusieurs cystéines et histidines sont conservées dans le domaine conservé C-terminal: cinq histidines et six cystéines sont présentes du LFY de Mousse *Physcomitrella patens* au LFY d'*A. thaliana* (Fig.29A). Or

ces résidus sont connus dans certains cas pour assurer la coordination d'ions métalliques tels le cobalt, le cuivre, le nickel, le fer, le magnésium ou encore le zinc, et de très nombreux facteurs de transcription coordonnent un ou plusieurs atomes de zinc comme démontré récemment pour les facteurs de transcription SBP par exemple (Fig.29B; Yamasaki et al., 2004). Nous avons donc souhaité étudier si tel était le cas pour le domaine C-terminal de LFY.

Les expériences développées ici nécessitent peu de protéine. Ainsi, sachant que le but ultime de notre projet est d'étudier la protéine LFY FL (LFY FL, *Full Length*), les expériences ont été menées en parallèle sur LFY-C, portion de la protéine qui contient les histidines et cystéines conservés, et sur LFY FL dont la production sera décrite dans le chapitre III.

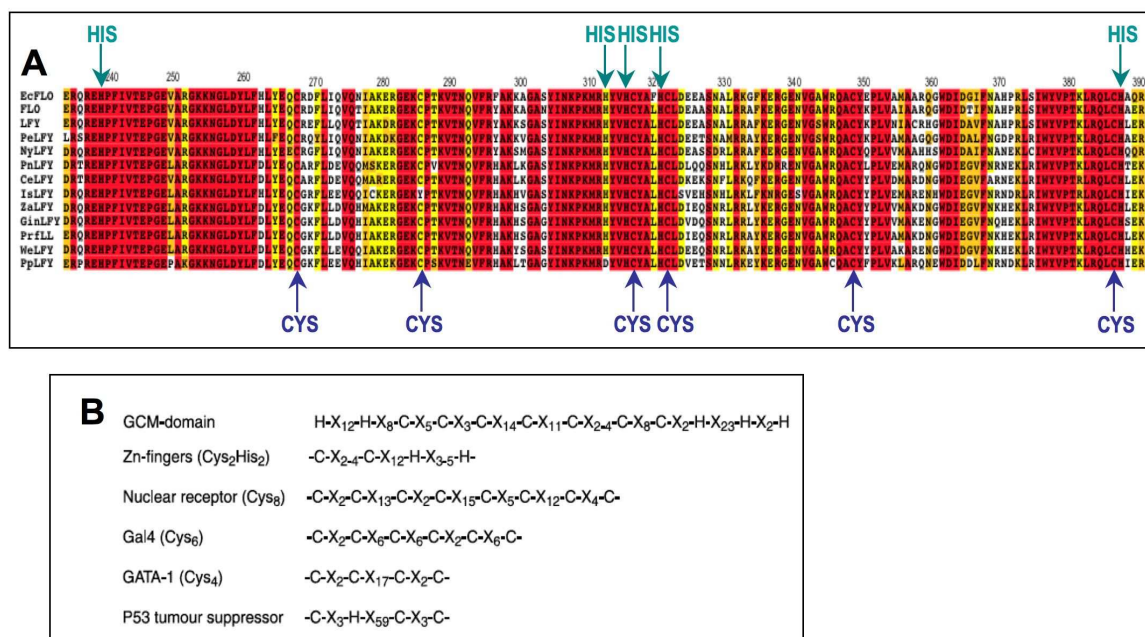


Figure 29: Les indices d'une coordination de métaux

(A) Alignement de séquences entre LFY-C (*A. thaliana*) et 12 protéines orthologues les Angiospermes eudicotylédones *E. californica* (EcFLO) et *A. majus* (FLO), les angiospermes basales *Peperomia sp.* (PeLFY) et *N. aquatica* (NyLFY), les Gymnospermes *Z. furfuracea* (ZaLFY), *G. biloba* (GinLFY), *P. radiata* (PRFL) et *W. mirabilis* (WeINDLY); les Fougères *P. nudum* (PnLFY), *C. richardii* (CeLFY) et *I. aquatica* (IsLFY), et la Mousse *P. patens* (PpLFY1). En rouge, résidus identiques; en orange, résidus aux propriétés biochimiques identiques; en jaune, résidus identiques à une exception près sur la présente sélection. Les histidines et cystéines conservées sont signalées; (B) Exemples de protéines coordinatrices de zinc et séquence consensus de coordination correspondante (Coen et al., 2002).

b) Résultats préliminaires par EMSA et AF

Pour étudier si une protéine coordonne un ou des ion(s) métallique(s), une stratégie classique consiste à étudier sa fonctionnalité suite à un traitement par un chélateur d'ions métalliques. Il existe de nombreux chélateurs dont le spectre d'action peut considérablement varier, certains étant spécifiques d'un ion métallique en particulier, d'autres pouvant chélater plusieurs ions. Etant donné que nous n'avons aucune idée de la nature du métal qui pourrait être coordonné par LFY, nous avons eu recours à un chélateur à large spectre d'action, l'EDTA. En parallèle, vu le grand nombre de facteurs de

transcription coordonnant du zinc (Jacobs, 1992), nous avons ciblé un chélateur fort, spécifique du zinc, à savoir la 1,10-phénantroline.

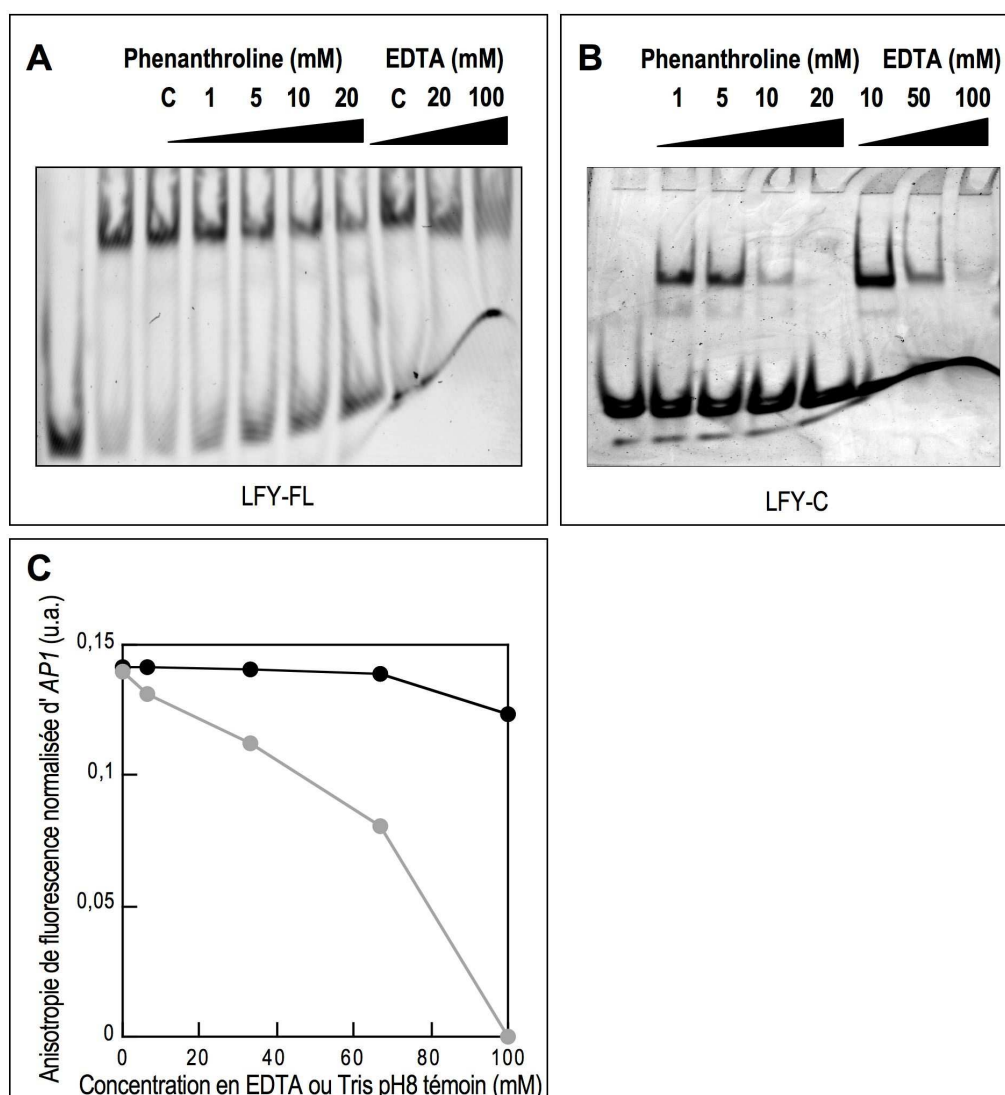


Figure 30: Etude de l'effet de chélateurs de métaux sur la fonctionnalité de LFY

(A) EMSA réalisé avec 10nM AP1 et 200nM LFY FL. Les concentrations et la nature des chélateurs sont précisées ; les contrôles (c) sont réalisés avec de l'éthanol dans le cas de la phénantroline et du Tris pH8.0 dans le cas de l'EDTA aux concentrations équivalentes à l'apport respectivement du traitement phénantroline et EDTA; (B) EMSA réalisé dans les mêmes conditions à partir de 200nM LFY-C ; (C) AF réalisé à partir de 10nM AP1, 160nM LFY FL, et une concentration croissante en EDTA (précisée en axe des abscisses). La courbe noire représente l'expérience contrôle mimant la dilution et l'apport en Tris pH8.0 de l'EDTA, la courbe grise représente l'expérience avec l'EDTA.

1-Conséquences d'un traitement simple EDTA ou phénantroline

L'effet d'un traitement EDTA ou phénantroline semble affecter l'affinité aussi bien de LFY FL que de LFY-C pour l'ADN, ce qui se traduit par une baisse d'AF ou une diminution d'intensité de la bande retardée en EMSA (Fig.30). Remarquons néanmoins que LFY-C semble plus sensible que LFY FL. Les résultats obtenus jusqu'ici, bien qu'encourageants, sont difficilement exploitables à eux seuls car ils sont réalisés en très large excès de chélateur par rapport à la protéine, excès qui pourrait être

rendu nécessaire en partie par une différence d'affinité entre ces chélateurs et LFY pour le métal, ou par une inaccessibilité de l'ion métallique lorsqu'il est associé à LFY.

2-Dénaturation à la chaleur et traitement phénantroline

De nombreuses protéines coordinatrices d'ions métalliques renferment ces ions en plein cœur de leur structure tridimensionnelle, les rendant difficilement accessible à tout chélateur extérieur (Nielsen et al., 2003; Whittaker et al., 2006). Pour tester si tel était le cas pour LFY, nous avons étudié si un cycle de dénaturation de LFY en présence du chélateur suivi d'une renaturation augmentait l'effet du chélateur. Mentionnons que nous nous sommes seulement intéressés ici à l'effet de la phénantroline dont le suivi en EMSA est plus aisé que celui de l'EDTA (l'EDTA perturbant visiblement la migration de l'ADN).

Pour dénaturer LFY, nous lui infligeons un traitement de 2min à 95°C suivi d'un retour immédiat à 4°C. Une grande partie de cette protéine dénaturée semble mal se replier, nécessitant de réaliser l'expérience avec une concentration 10 fois plus élevée en protéines (Fig.31). Mais de façon intéressante, la part qui parvient tout de même à se renaturer semble plus sensible à la phénantroline: alors qu'il était nécessaire de traiter les protéines à 20mM pour pouvoir voir un effet significatif chez les protéines non dénaturées, une concentration de 5mM suffit ici.

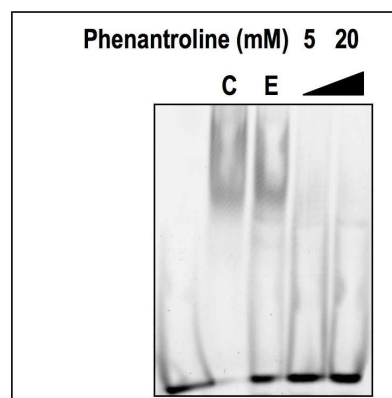


Figure 31: Etude de l'effet de la phénantroline sur LFY dénaturée à la chaleur

EMSA réalisé avec 10nM AP1 et 2000nM LFY FL dénaturée à la chaleur et renaturée. La présence et concentration en phénantroline est signalée sur la figure. C = protéine seule, E = protéine + éthanol reproduisant l'apport en éthanol occasionné par l'ajout de phénantroline.

Nous nous sommes ensuite intéressés à la capacité d'ions métalliques à restaurer la perte de liaison résultant du traitement phénantroline. Le principe de cette expérience est de tester quels ions métalliques permettent de renaturer LFY après un traitement phénantroline, et donc de connaître la nature des ions coordonnés par LFY. Plusieurs ions ont été testés à ces fins: le zinc, le nickel, le manganèse, le chlore, le cuivre et le fer. De ces différents ions un seul a entraîné un effet visible sur la capacité de LFY à lier l'ADN, le zinc (Fig.32). LFY semble donc coordonner un ou des ion(s) zinc.

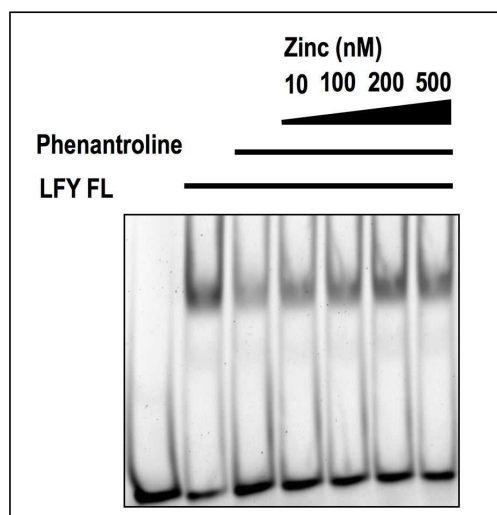


Figure 32: Essai de complémentation de l'effet phénantroline par le zinc

EMSA réalisé avec 10nM AP1, 120nM LFY FL, 20mM phénantroline, et une gamme de concentration en $ZnCl_2$.

3-Bilan des résultats préliminaires

Ces résultats suggèrent que LFY pourrait coordonner un ou des ion(s) métallique(s), ion(s) zinc vraisemblablement.

Je n'ai pas abordé tout au long de ce premier descriptif la grande difficulté de reproductibilité que j'ai rencontré tout au long de ces manipulations, difficulté particulièrement prononcée dans le cas de l'expérience de dénaturation/renaturation. Il est donc devenu nécessaire de faire appel à d'autres méthodologies plus sensibles pour tester si LFY est effectivement un coordinateur de métaux, et si tel est le cas, pour déterminer le ratio protéine/ion(s) métallique(s).

c) Analyses fines par dosage des métaux

Pour cela, nous avons eu recours à deux techniques indépendantes de dosage d'ions métalliques présents en trace à partir d'une préparation de LFY en solution. L'ICP-MS a été réalisée en collaboration avec Corinne Rivasseau, chercheur au sein de notre laboratoire, tandis que le microPIXE est le fruit d'une collaboration avec Elspeth Garman (Université d'Oxford, Royaume-Uni).

1-ICP-MS (*Inductively Coupled Plasma Mass Spectrometry*)

L'ICP-MS est une technique de spectrométrie de masse couplée à un plasma inductif. Cette technique d'analyse repose sur la séparation, l'identification et la quantification des éléments constitutifs d'un échantillon en fonction de leur masse. Elle est basée sur le couplage d'une torche à plasma générant des ions à partir de l'échantillon et d'un spectromètre de masse (Thermoelectron X7 dans le cas de l'ICP-MS) qui sépare ces ions en fonction de leur masse. Pour cela, cette technique nécessite une étape préalable de minéralisation de l'échantillon à partir d'un mélange d'acide nitrique et d'acide chlorhydrique chauffé à 200°C.

La protéine LFY FL a été analysée à deux reprises par cette technique. Lors du premier essai, nous avons dosé quatre éléments, à savoir le fer, le zinc, le magnésium et le cuivre. L'absence de fer et cuivre a pu être affirmée dès ce premier essai. En revanche, la détection de zinc et de magnésium à un ratio proche de 1 atome/protéine a nécessité une deuxième analyse. Cette dernière confirmera que les éléments zinc et magnésium sont présents en une concentration inférieure à un ratio 1/1 par rapport à celle de LFY, réfutant également l'hypothèse de la coordination de zinc et de magnésium par LFY.

2-microPIXE

L'analyse des éléments d'une protéine par technique de microPIXE repose sur l'étude de l'émission rayons X de particules induites par un microfaisceau de protons, émission caractéristique de chaque élément (Garman and Grime, 2005). Elle présente un avantage certain par rapport à l'ICP-MS, à savoir que le dosage des atomes de soufre présents dans les méthionines et les cystéines de la protéine apporte une calibration interne de la quantité de protéines présente. Nous avons donc demandé à Elspeth Garman d'analyser un échantillon de LFY-C et à son tour, elle nous a confirmé que LFY-C ne contenait pas d'ion métallique (moins de 0,1 atome lourd par molécule de LFY).

Il semble donc finalement exclu que LFY coordonne un ou des ion(s) métalliques.

8. Discussion

a) Ce que l'approche biochimique nous a apporté

1-Caractérisation de l'interaction LFY-C/ADN

Le domaine C-terminal de LFY est un domaine fonctionnel suffisant à la liaison à l'ADN. Nous proposons ici un affinement de la délimitation du domaine minimal décrite par Maizel et collaborateurs (2005) aux acides aminés 230-391: la région hydrophobe C-terminale n'intervient donc pas dans le contact à l'ADN. La protéine produite à partir du vecteur pCH42 (résidus 241-421), capable de lier l'ADN mais avec une affinité fortement affectée, ce qui suggère qu'un ou plusieurs résidus compris dans la région 230-241 joueraient un rôle important dans cette interaction.

LFY-C lie spécifiquement l'ADN selon une affinité en accord avec les gammes d'affinité décrites dans la littérature, soulignant la réalité biologique de cette liaison. Mentionnons que les K_d indiqués ici sont des K_d apparents avec incertitude sur la mesure, ne prenant pas en compte notamment le degré d'oligomérisation de LFY sur l'ADN.

Par AF, nous avons constaté que cette affinité pouvait être perturbée par une diminution de la taille de l'oligonucléotide testé, suggérant que LFY contacte l'ADN sur au moins 22pb.

La technique de SPR a permis d'estimer les affinités de LFY pour les régions cis-régulatrices entières de trois de ses gènes cibles, *AP1*, *AP3* et *AG*. En prenant le cas d'*AP1*, l'affinité estimée, de 15nM, est plus élevée que celle de LFY pour l'oligonucléotide *AP1* estimée par AF et EMSA à 100nM. Peut-on mettre en cause des variations inhérentes aux techniques utilisées ? Pour le savoir il faudrait

réaliser une expérience de SPR dans les mêmes conditions qu'en AF ou EMSA, c'est à dire avec un oligonucléotide de 30pb. Mais il se pourrait aussi que cet écart vienne d'une différence de comportement de LFY sur l'oligonucléotide ou le promoteur *AP1*. En effet, nous y reviendrons dans le chapitre III, mais il se pourrait que LFY ait d'autres sites de liaisons au niveau du promoteur d'*AP1*.

2-Mise en évidence d'une coopérativité LFY-C/LFY-C sur l'ADN

LFY-C seule en solution est majoritairement monomérique. En revanche, en présence d'ADN, deux complexes au ratio LFY-C/ADN différent sont observés. Le plus petit complexe (C1) contient une protéine LFY-C, tandis que le complexe le plus important (C2) en contient deux. Le fait que le complexe majoritaire soit systématiquement ce complexe C2, indépendamment de la concentration en LFY-C, est le résultat d'un mécanisme de coopérativité.

Nous avons vu que cette coopérativité est dépendante de l'ADN: elle est davantage affectée par la mutation *AP1m1* qu'*AP1m2*, ce qui suggère que chaque monomère contacterait un demi-palindrome avec une affinité différente.

Les protéines produites à partir du vecteur pCH43 (résidus 224-391; Fig.23) semblent avoir une coopérativité comparable à celle des protéines produites à partir de pCH28 où le complexe C2 est toujours majoritaire devant le complexe C1 indépendamment de la concentration protéique. Ces données suggèrent donc que la queue hydrophobe non conservée à l'extrémité C-terminale de LFY n'interviendrait pas dans ce mécanisme.

En revanche, même si les protéines produites à partir du vecteur pCH42 (résidus 241-421; Fig.23) sont toujours capables de lier l'ADN sous la forme des complexes C1 et C2, elles semblent avoir une coopérativité affectée: la bande retardée correspondante au complexe C1 est d'intensité comparable à celle du complexe C2 (EMSA à 1000, 2000 et 3000nM de protéines; Fig.23). Cette observation pourrait suggérer que la région 230-241 comporte des résidus impliqués dans cette coopérativité, ou dans le bon repliement de la protéine.

Il est important de signaler que ces résultats ne permettent pas de définir quel est le mécanisme moléculaire à la base de cette coopérativité de liaison. En effet, rien ne permet jusqu'à présent d'affirmer s'il résulte d'un changement de conformation de l'ADN suite à la liaison du premier monomère, ou à une dimérisation de LFY. Peut-on vraiment parler ici de dimérisation ? Rien ne permet d'affirmer jusqu'ici que les deux monomères de LFY interagissent physiquement. Seule l'étude structurale du complexe LFY-C/ADN permettra de répondre à cette question.

3-Une hypothèse rejetée: le domaine C-terminal de LFY ne coordonne pas d'ions métalliques

L'alignement de séquence entre orthologues de LFY chez différentes espèces végétales révèle l'intrigante conservation de résidus cystéines et histidines qui nous a porté à croire que le domaine C-terminal de LFY pourrait coordonner un ou des ion(s) métallique(s). Cette hypothèse a été renforcée

par une série de résultats préliminaires basés sur l'étude de l'effet de chélateurs de métaux sur la capacité de LFY à lier l'ADN. Pourtant, deux expériences indépendantes de dosage de métaux réfutent définitivement cette hypothèse.

b) Ce à quoi l'approche biochimique ne permet pas de répondre

Même si nos résultats délimitent un domaine minimal de liaison à l'ADN, ils ne nous permettent pas de définir comment est organisé le DBD: combien de résidus de LFY-C sont impliqués dans cette reconnaissance ? Où sont-ils localisés? Existe-t-il une région monofonctionnelle de liaison à l'ADN, ou ces résidus sont-ils dispersés tout au long du domaine minimal ?

Nous savons désormais que deux unités de LFY peuvent se lier simultanément à l'ADN. Mais comment chaque monomère se positionne-t-il l'un par rapport à l'autre sur ce brin d'ADN? Interagissent-ils physiquement l'un avec l'autre? Cette question revient à identifier quel mécanisme moléculaire est à l'origine de cette coopérativité: est-ce le résultat d'une modification conformationnelle de l'ADN induite par C1, ou est ce le résultat d'une véritable dimérisation de LFY ? Seule l'étude structurale du complexe LFY-C/ADN permettra de répondre à ces questions.

c) Réflexions sur nos résultats

1-Les outils méthodologiques:

L'AF présente différents avantages: extrêmement facile à mettre en œuvre, elle permet de détecter des interactions en solution ce qui limite toute perturbation des équilibres réactionnels. Elle délivre les résultats en temps réel, ce qui donne potentiellement accès à toutes les constantes cinétiques. Dans le cas présent, nous n'avons pas étudié les constantes cinétiques d'association et de dissociation mais les mesures d'AF de LFY-C/AP1, réalisées sans temps d'incubation, suggèrent une cinétique d'association rapide et/ou de dissociation lente. Toutefois, malgré leurs avantages, les mesures en solution présentent également un inconvénient incontestable: elles ne permettent pas d'avoir accès aux différentes espèces en présence. Dans le cas d'une courbe de saturation typique de LFY-C en présence d'AP1, il n'est pas possible de distinguer le complexe C1 du complexe C2: l'AF donne une vision globale de la réaction, et permet donc d'estimer une affinité apparente résultante de l'affinité du complexe C1 et du complexe C2 pour l'ADN.

L'EMSA au contraire permet de visualiser trois des quatre espèces présentes dans la réaction de liaison de LFY-C à l'ADN, à savoir l'ADN libre, le complexe C1 et le complexe C2. Seul manque protéine libre dont nous pourrions cependant suivre la présence si nécessaire à l'aide d'un marqueur fluorescent. Ici l'inconvénient est l'utilisation d'un support, le gel d'acrylamide, et la nécessité de laisser migrer les espèces en présence sur un temps non négligeable (1h minimum), deux paramètres qui peuvent influencer les équilibres réactionnels.

Pouvoir travailler en AF et EMSA permet donc de s'affranchir des inconvénients inhérents à chacune de ces deux techniques, et constitue un outil indispensable à la caractérisation de liaison de facteurs de transcription à leur ADN cible.

La technique de SPR, simple, rapide et quantitative, offre quant à elle d'intéressantes perspectives pour l'identification de nouvelles cibles directes d'ADN de facteurs de transcription. En testant la capacité de liaison entre facteur de transcription et régions cis-régulatrices entières d'un gène donné, cette technique permet de travailler 'à l'aveugle' sans avoir idée de la séquence consensus de fixation du facteur de transcription.

2-Les limites rencontrées

La première série de résultats visant à identifier la nature des complexes C1 et C2 soulève la difficulté résidente dans la détermination du poids moléculaire d'un complexe en conditions non dénaturantes. Ces difficultés s'expliquent probablement dans le fait que les techniques utilisées ici ne sont pas uniquement fonction de la masse de la molécule considérée, mais également de sa conformation et/ou de sa charge. De plus, elles sollicitent une matrice qui pourrait ne pas être neutre sur les propriétés du complexe. Les stratégies de MALLS-SEC, de GEMSA et d'EMSA en présence de deux protéines de taille différente représentent une alternative efficace pour contourner ces difficultés.

Comment expliquer les effets de la phénantroline et de l'EDTA sur la liaison de LFY-C à l'ADN ? Nous avons déjà mentionné que ces chélateurs ont été utilisés à très forte concentration. Ainsi, il serait possible qu'ils viennent perturber la conformation de la protéine ou de l'ADN par la chélation indirecte d'ions du milieu environnant, ions tels le magnésium connus pour intervenir dans l'organisation du double brin d'ADN (Sines et al., 2000). Dans le cas de l'expérience de dénaturation/renaturation de LFY, le fait que la phénantroline agisse à des concentrations plus faibles pourrait s'expliquer par une fragilité de la protéine conséquence au traitement à la chaleur, fragilité qui la rend plus sensible à l'effet indirect de ce chélateur. Il aurait pu être bénéfique d'utiliser en complément des contrôles de liaison avec d'autres facteurs de transcription ne coordonnant pas d'ions métalliques.

II. Etude structurale du domaine conservé C-terminal de LFY

L'analyse de LFY-C par approche biochimique classique a permis d'améliorer notre compréhension de son mode de liaison à l'ADN. Toutefois, cette approche ne permet pas d'identifier les résidus clés impliqués de ce processus, qu'il s'agisse des résidus nécessaires au contact à l'ADN et à l'éventuelle dimérisation de LFY-C. C'est pourquoi nous avons entrepris une étude structurale du complexe LFY-C/ADN.

La première étape critique à la réalisation d'une étude cristallographique est la production de la protéine d'intérêt en forte concentration. Dans cette optique, j'ai consacré une large partie de ma première année de thèse à essayer de produire la protéine LFY FL, mais les quantités obtenues se sont révélées toujours insuffisantes et n'ont permis d'aborder qu'une caractérisation biochimique de cette protéine, nous y reviendrons dans le chapitre III. Contrairement à la protéine LFY FL, la protéine LFY-C est très bien exprimée sans nécessité d'apporter une étiquette de solubilisation, se purifie selon un protocole simple et extrêmement reproductible, supporte le clivage de son étiquette 6his, et supporte d'être concentrée à un degré permettant la cristallogenèse.

1. Protocole de cristallogenèse

La cristallogenèse est la deuxième étape critique au processus de cristallographie. Elle est extrêmement exigeante quant à la qualité de la molécule ou du complexe à cristalliser. Le caractère homogène ou monodispersité du complexe LFY-C/AP1 a été observée par microscopie électronique (en collaboration avec Guy Schoen, EMBL Grenoble) et par DLS (Dynamic Light Scattering; Fig.33).

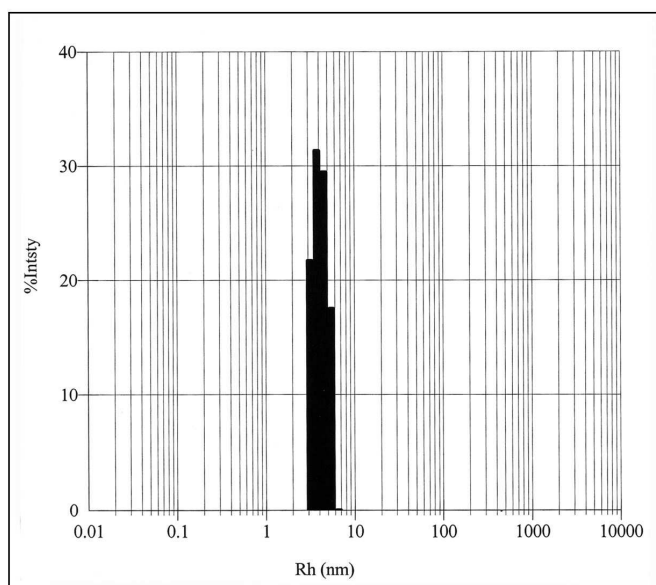


Figure 33: Etude de la monodispersité du complexe LFY-C/ADN à cristalliser par DLS

La cristallogenèse est également imprévisible. Les conditions de cristallogenèse, fonction du pH, de la nature et de la concentration du précipitant, du sel et d'éventuels additifs, ne suivent aucune loi,

sont propres à chaque molécule ou complexe. Dans le cadre de la cristallogénèse d'un complexe protéine/ADN, plusieurs données suggèrent que les choix de la construction protéique et de l'oligonucléotide cible sont déterminants. C'est à la lumière de ces connaissances que nous avons commencé à dessiner des oligonucléotides de nature (*AP1*, *AG-I*, *AG-II*), de taille et extrémités (franches ou cohésives) différentes (Récapitulatif dans le tableau 2). Le crible de cristallogénèse a donc pris en compte deux paramètres: la nature du complexe à cristalliser, et la nature des conditions de cristallisation.

Tableau 2: Bilan des différents oligonucléotides testés en cristallogénèse

Seul le brin sens (5'-3') est représenté. La base médiane de la séquence consensus CCANTG est précisée en bleu. Les extrémités dépassantes sont différenciées en extrémités cohésives (permettant de créer des liaisons covalentes entre oligonucléotides voisins) et non cohésives. La base dépassante est précisée en vert. Les bases modifiées de l'oligonucléotide *AP1*-20pb symétrique construit pour avoir deux demi-sites équivalents sont représentées en rouge.

Oligonucléotide	Nombre de bases	Extrémités	Séquence
<i>AP1</i>	20 symétrique	Franches	TACGGACCAGTGGTCCGTAC
<i>AP1</i>	20	Franches	GAAGGACCAGTGGTCCGTAC
<i>AP1</i>	21	Franches	GGAAGGACCAGTGGTCCGTAC
<i>AP1</i>	22	Franches	GGAAGGACCAGTGGTCCGTACA
<i>AP1</i>	23	Franches	GGGAAGGACCAGTGGTCCGTACA
<i>AP1</i>	24	Franches	GGGAAGGACCAGTGGTCCGTACAA
<i>AP1</i>	25	Franches	GGGAAGGACCAGTGGTCCGTACAA
<i>AP1</i>	19	Cohésives	AAAGGACCAGTGGTCCGTA
<i>AP1</i>	20	Cohésives	AGAAGGACCAGTGGTCCGTA
<i>AP1</i>	21	Cohésives	AGAAGGACCAGTGGTCCGTAC
<i>AP1</i>	22	Non cohésives	GGGAAGGACCAGTGGTCCGTAC
<i>AP1</i>	23	Cohésives	AGGAAGGACCAGTGGTCCGTACA
<i>AP1</i>	24	Cohésives	AGGAAGGACCAGTGGTCCGTACA
<i>AP1</i>	25	Cohésives	AGGAAGGACCAGTGGTCCGTACAA
<i>AP1</i>	26	Non cohésives	TGGGAAGGACCAGTGGTCCGTACAA
<i>AG-I</i>	20	Cohésives	ATTTAATCCAATGGTTACAA
<i>AG-I</i>	21	Cohésives	AATTTAATCCAATGGTTACAA
<i>AG-II</i>	20	Cohésives	ATTATACCCAATGTGTTAAT
<i>AG-II</i>	21	Cohésives	ATTATACCCAATGTGTTAAT

Initialement, nous avons essayé de cristalliser le complexe LFY-C/*AP1* à partir de deux oligonucléotides différents (*AP1*-21pb extrémités franches et *AP1*-22pb extrémités dépassantes; Tab.2) en système automatisé et à température ambiante à partir d'un crible de 48 conditions (robot de cristallogénèse de l'EMBL Grenoble). La grande majorité des gouttes réalisées ayant formé des précipités, nous avons poursuivi la cristallogénèse à 4°C, manuellement cette fois ci, à partir d'un crible commercial de 48 conditions établi à partir des principales conditions de cristallogénèse de complexe protéine/DNA connues (kit Natrix, Hampton). Une partie des cribles a été réalisée selon un ratio ADN:protéine 1:4, ratio qui a été rectifié à 1:2 suite aux résultats de MALLS-SEC. Ce changement de

ratio a conduit à une diminution du pourcentage de gouttes précipitées. Le tableau 3 résume les conditions ayant conduit à la formation de cristaux analysés au synchrotron.

Tableau 3: Résumé des différentes conditions de cristallogenèse ayant abouti à la formation de cristaux LFY-C/ADN analysés au synchrotron

Oligonucléotide	Construction protéique	Ratio protéine:ADN	pH	Nature du précipitant	Taille des cristaux [μm^3]	Forme du cristal	Pouvoir de diffraction
AP1 25mer 1bp dépassante	pCH28	4:1	6,5	PEG4k	200x10x10	Aiguilles	>6 Å
AP1 21mer extrémités franches	pCH28	4:1 2:1	7,0 7,0	PEG4k PEG400	150x10x10 200x30x10	Aiguilles Plaque 2-D	>6 Å >6 Å
AP1 20mer 1bp T:A dépassante	pCH28	4:1 4:1 4:1 2:1	6,0 6,5 8,5 6,0	Isopropanol PEG4K PEG400 Isopropanol	150x20x10	Plaque 2-D Aiguilles Polyèdre Parallélépipède	4 Å >6 Å >6 Å 4Å
AP1 20mer extrémités franches	pCH26	2:1	8,5	PEG400	20x20x150	Bâtonnet	3Å
AP1 20mer 1bp T:A dépassante	pCH26	2:1	7,0	PEG400	300x300x500	Bipyramide	2.1Å
AG-I 20mer 1bp T:A dépassante	pCH26	2:1	7,0	PEG400	300x300x500	Bipyramide	2.2Å

2. Cristallographie

La résolution de la structure tridimensionnelle de LFY-C, résidus 237 à 399, lié à AP1 et à AG-I a été réalisée intégralement par nos collaborateurs. Le phasage a été effectué par trempage des cristaux dans une solution de mercure. Pour plus de renseignements à ce sujet, se référer à l'article Hamès et collaborateurs (Annexe 1).

a) Structure globale

Malgré la différence des motifs de reconnaissance de LFY au niveau d'AP1 et AG-I (Fig.12 et tab.2), les deux structures obtenues se superposent largement. La structure globale du complexe est formée d'un dimère de LFY-C contactant l'ADN. Chaque monomère contacte principalement un demi-site (délimité par la base médiane soulignée en fig.12) à l'exception de la Lys284 qui contacte le squelette phosphate reliant les guanines invariantes aux positions ± 2 et ± 3 du demi-palindrome opposé. LFY-C adopte une structure compacte globulaire constituée de deux courts feuillets β suivis de sept hélices α connectées par des petites boucles (Fig.34A et B). L'absence de larges régions hydrophobes à l'extérieur de la structure suggère que LFY-C représente un domaine fonctionnel

autonome de liaison à l'ADN, en accord avec une action indépendante au domaine conservé N-terminal comme le suggéraient les analyses de truncations de LFY (Fig.23; Maizel et al., 2005).

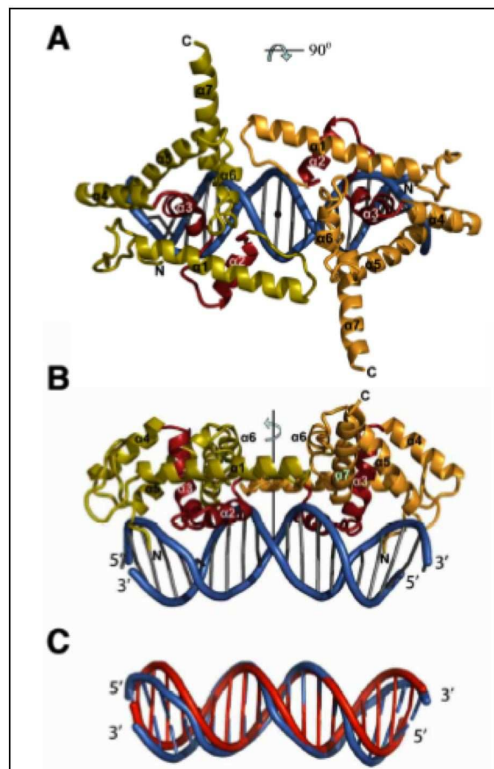


Figure 34: Structure du dimère de LFY-C lié à AP1

(A-B) Deux vues orthogonales du dimère de LFY-C (résidus 237-399) lié à l'ADN. Les monomères sont colorés en orange et olive et leur motif HTH (hélices $\alpha 2$ et $\alpha 3$) en rouge. L'ADN est représenté en bleu; (C) Superposition de l'ADN double brin du complexe LFY-C/ADN (en bleu) à la forme B régulière de l'ADN (en rouge).

b) Caractéristiques du domaine de liaison à l'ADN de LFY-C

Le domaine de liaison de LFY-C à l'ADN (LFY-DBD) est défini par deux motifs: un motif contactant le grand sillon de l'ADN mettant en jeu les hélices $\alpha 2$ et $\alpha 3$ qui forment un motif HTH où $\alpha 3$ s'oriente parallèlement au brin d'ADN pour ne le contacter que par son extrémité N-terminale (Fig.34A et B; Aravind et al., 2005); et un motif représenté par l'Arg237 qui forme un long prolongement venant s'insérer dans le petit sillon de l'ADN (Fig.35). La structure de l'ADN montre une courbure de 20° qui permettrait un rapprochement du petit sillon au grand sillon (Fig.34C).

Chaque hélice ($\alpha 2$ et $\alpha 3$) compte un point de contact base-spécifique: l'Asn291 et la Lys307 reconnaissent les deux guanines invariantes aux positions ± 2 et ± 3 (Fig.35). L'Arg237, premier résidu ordonné en position N-terminale de la structure, réalise le troisième contact base-spécifique au niveau de la paire de base A:T ± 8 (Fig.35). Si les deux premières bases reconnues par LFY-DBD étaient déjà identifiées dans la séquence consensus (correspondant aux bases opposées aux deux cytosines juxtaposées de la séquence CCANTG^{G/T}), cette base en position ± 8 était totalement ignorée. Pour cause, cette base A:T chez AP1 est inversée en T:A chez AG-I. La comparaison des deux structures

LFY-C/AP1 et LFY-C/AG-I révèle que l'Arg237 forme un 'bras' souple qui dans un cas contacte le brin sens, et dans l'autre cas le brin antisens (Fig.36). L'importance de cette position est soulignée par la présence systématique d'une base A:T ou T:A dans tous les sites confirmés (séquences AP1, AG-I, AG-II et AP3-I; Hong et al., 2003; Lamb et al., 2002; Parcy et al., 1998). Ces données structurales apportent donc une donnée supplémentaire au site consensus de LFY qui peut désormais être plus précisément défini par la séquence $A/TNNNNCCANTGT/GNNNNNT/A$.

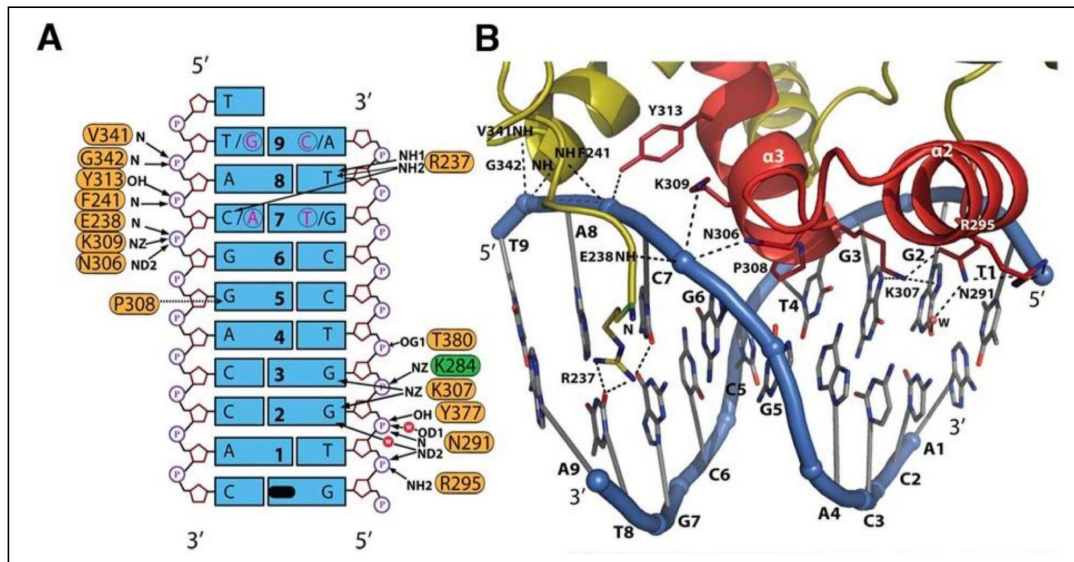


Figure 35: Reconnaissance d'AP1 par LFY-C

(A) Détails des interactions LFY-C/ADN sur un demi-site AP1. L'ADN est schématisé en bleu, les résidus du premier monomère en orange. Les paires de base 7 et 9 correspondant au second demi-site sont représentées en rose et entourées. Les interactions polaires et hydrophobes sont représentées respectivement par des flèches solides ou en pointillés. La Lys284 représentée en vert appartient au monomère contactant principalement le second demi-site. La base médiane est marquée d'un trait noir; (B) Représentation schématique d'un monomère de LFY-C lié à un demi-site AP1. La protéine est représentée en olive à l'exception de son motif HTH en rouge. Les interactions polaires sont ici représentées en pointillés.

En plus de ces trois contacts spécifiques, LFY-DBD compte douze résidus contactant l'ADN au niveau du squelette phosphate, résidus situés pour la plupart dans l'hélice $\alpha 3$ (Fig.35). En particulier, un point de contact inhabituel est réalisé par la Pro308 qui exerce une pression physique ('stacking') entre les deux guanines situées en position ± 5 et ± 6 .

c) Similarités structurales

Les hélices $\alpha 1$, $\alpha 2$ et $\alpha 3$ de LFY-C présentent une importante homologie structurale avec des motifs HTH tels ceux de la protéine à homéodomaine engrailed, du domaine paired et de la tc3 transposase (Fraenkel et al., 1998; van Pouderoyen et al., 1997; Xu et al., 1999). Cependant, LFY-DBD peut difficilement être assigné à ces classes de protéines. L'orientation de son hélice $\alpha 3$ diffère de celle des protéines à homéodomaine qui s'aplatit davantage sur le grand sillon de l'ADN pour le contacter en son milieu (Fig.37). En revanche, en plus d'une orientation similaire de l'hélice $\alpha 3$ de LFY-C et du

domaine paired (Fig.37), ces deux domaines possèdent deux courts feuillets β . La différence entre leur DBD vient du point de contact au petit sillon: dans le cas du domaine paired, il est réalisé par la boucle positionnée entre les deux courts feuillets β , alors que dans le cas de LFY-C, il est réalisé par l'Arg237 à l'extrémité N-terminale. Le domaine C-terminal de LFY représenterait donc un nouveau variant de cette grande famille que représentent les protéines HTH.

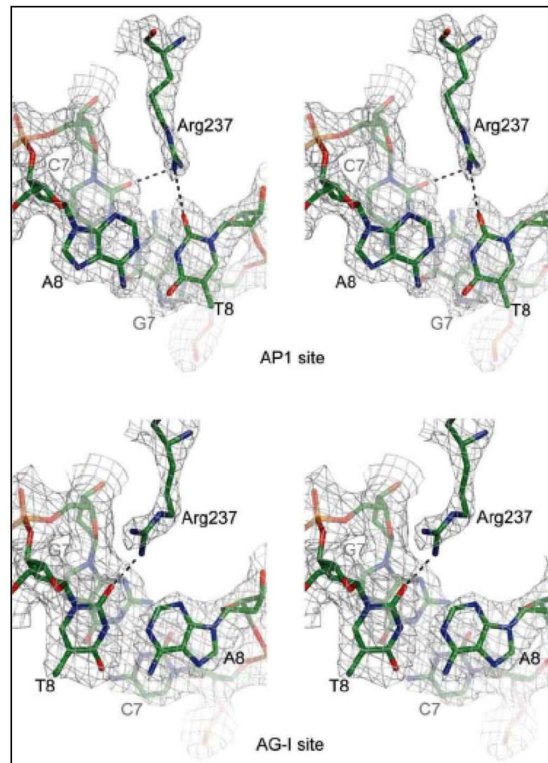


Figure 36: Carte de densité électronique de l'interaction Arg237/ADN

L'Arg237 contacte AP1 au niveau des bases ± 7 et ± 8 (en haut) et AG-I au niveau de la base ± 8 du brin opposé (en bas).

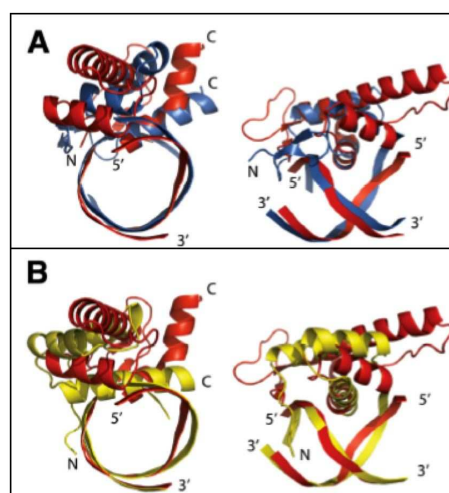


Figure 37: Comparaison structurale du DBD de LFY-C au domaine paired et à l'homéodomaine

Deux vues orthogonales des hélices $\alpha 1$, $\alpha 2$ et $\alpha 3$ de LFY-C liées à leur séquence cible AP1 (en rouge) superposées (A) aux 3 hélices de l'extrémité N-terminale du domaine paired Prd de Drosophile (en bleu, PDB-id: 1pdn); (B) à l'homéodomaine de la protéine engrailed de Drosophile lié à l'ADN (en jaune, PDB-id: 1hdd), où l'hélice $\alpha 3$ reconnaît le grand sillon de l'ADN en son milieu.

d) Vers la compréhension du mécanisme moléculaire de la coopérativité

La structure du complexe de LFY-C à l'ADN révèle également d'importantes interactions entre les deux monomères: l'Asp280 de la boucle comprise entre les hélices $\alpha 1$ et $\alpha 2$ est contacté au niveau de son squelette carboné par l'His387 et l'Arg390 de l'hélice $\alpha 7$ (Fig.38). Nous avons vu que la coopérativité pouvait résulter soit d'une modification de la conformation de l'ADN, soit d'une interaction monomère-monomère. La structure suggère donc que le mécanisme moléculaire à l'origine de la coopérativité pourrait être la dimérisation de LFY-C.

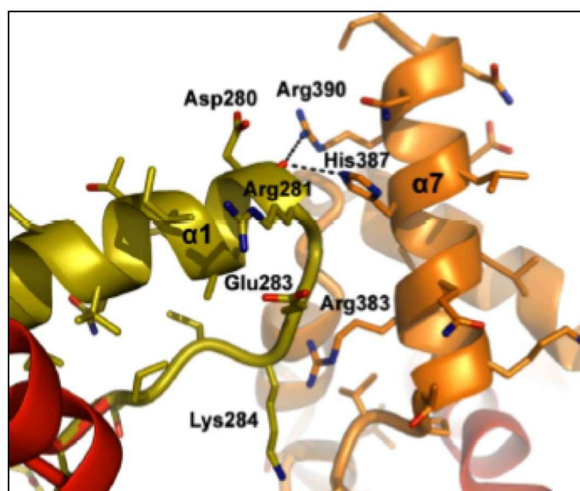


Figure 38: Détail des interactions LFY-C/LFY-C

Le premier monomère est représenté en olive (et son motif HTH en rouge), le second en orange. Les liaisons polaires sont représentées en pointillés.

3. Validation et approfondissement des données structurales par approche biochimique

a) Validation des points de contact LFY-C/AP1

1- par l'étude du comportement de liaison de protéines LFY-C mutées

Suite à la caractérisation structurale de LFY-DBD, nous avons souhaité étudier *in vitro* quelles étaient les conséquences de mutations ponctuelles en alanine (résidu neutre de faible encombrement moléculaire) des résidus de LFY-C jouant un rôle clé dans la reconnaissance de l'ADN. Quatre mutants protéiques ont ainsi été produits dans ce but: trois touchant les résidus impliqués dans la reconnaissance spécifique, à savoir l'Arg237, l'Asn291 et la Lys307, et un concernant la Pro308 responsable d'un *stacking*. En EMSA, les protéines R237A, N291A et K307A conduisent toutes les trois à une perte d'affinité de LFY-C pour AP1 comme l'illustre la figure 39 (A). En revanche, P308A ne semble pas modifier cette affinité car P308A et LFY-C parviennent à lier AP1 dès une concentration initiale identique et dans une proportion comparable (Fig.39A). Néanmoins, P308A affiche dès 750nM un comportement de liaison le différenciant de LFY-C, marqué par l'apparition d'une bande supplémentaire correspondant à un complexe LFY-C/ADN supérieur au ratio 2:1, complexe qui n'est observé avec LFY-C qu'à partir de 5000nM (Fig.39A-B).

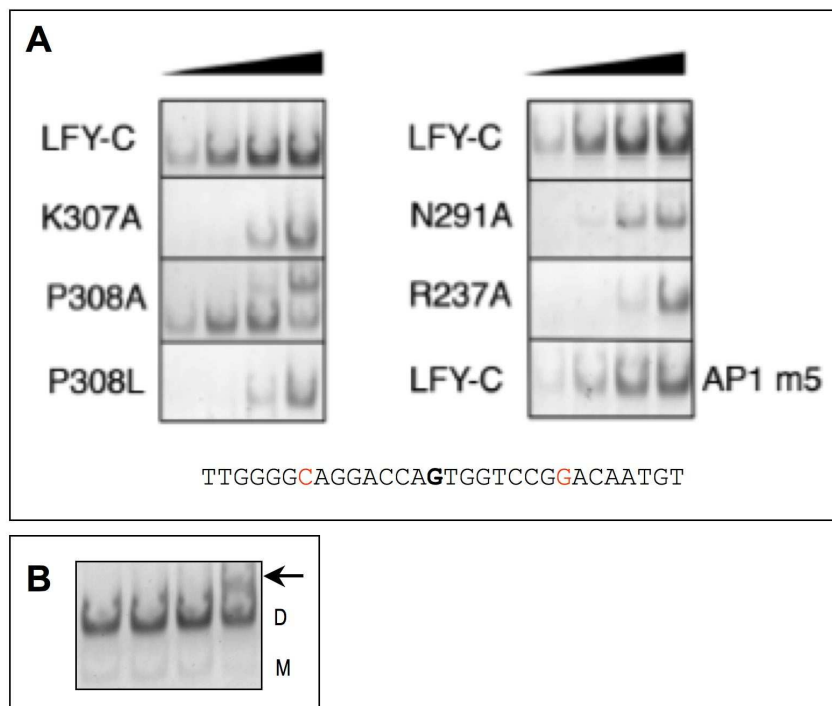


Figure 39: Validation des points de contact LFY-C/AP1 par EMSA

(A) Etude de l'effet des mutations de spécificité de reconnaissance de l'ADN sur l'affinité de liaison de LFY-C à AP1. EMSA réalisés avec 10nM AP1 et des concentrations croissantes en protéines (100, 250, 750, 2000nM de gauche à droite). Seul le complexe dimerique C2 est représenté si ce n'est dans le cas de P308A qui laisse apparaître un complexe supplémentaire d'ordre supérieur. La séquence d'AP1m5 est précisée en dessous des gels natifs, les mutations sont rapportées en rouge, la base médiane est indiquée en gras; (B) EMSA réalisé avec 10nM AP1 et une gamme de concentration en LFY-C (500, 1000, 2000 et 5000nM). Mise en évidence d'un complexe d'ordre supérieur à 5000nM LFY-C indiqué par une flèche noire. Les bandes correspondantes au dimère 'D' et au monomère 'M' sont signalées.

2- par l'étude de l'influence de mutations dans la séquence AP1

Nous avons vu que les bases reconnues spécifiquement par LFY-DBD étaient les bases guanine $\pm 2 \pm 3$ et la base thymine ± 8 . Pour valider ces résultats, nous avons étudié l'effet de mutations de ces bases sur l'affinité de LFY-C pour l'oligonucléotide AP1 et avons constaté qu'effectivement, les mutations simultanées des deux guanines (AP1m3) et la mutation de la thymine (AP1m5) entraînent une diminution d'affinité de LFY-C pour l'ADN par rapport à ce qui avait été observé avec la séquence AP1 sauvage (AP1 WT; Fig.28C et 39A). Notons cependant que l'affinité de LFY-C pour AP1m5 est nettement moins réduite que celle de R237A (résidu impliqué dans la reconnaissance de la base mutée en AP1m5) pour AP1 WT, signifiant que la mutation du résidu impliqué dans la reconnaissance de cette base thymine ± 8 est plus pénalisante que la mutation de la base d'ADN. Ce constat peut trouver explication dans le fait que l'Arg237 contacte également la base adjacente à la thymine mutée ± 8 (Fig.36): l'effet de la mutation de la base thymine (AP1m5) est amoindrie par la présence de cette interaction entre l'Arg237 et la base adjacente.

b) Etude du comportement de liaison des mutants connus de LFY *in planta*

Parmi les mutations ponctuelles décrites *in planta* dans la littérature, quatre sont situées à proximité d'un point de contact avec l'ADN: *lfy-5* (P240L) et *lfy-4* (E238K) sont localisées à proximité de l'Arg237, *lfy-20* (N306D) est adjacente à la Lys307, et *lfy-28* (P308L) correspond à la mutation de la Pro308. *In planta*, *lfy-4*, *lfy-5* et *lfy-20* présentent des phénotypes *lfy* faibles, tandis que *lfy-28* présente un phénotype intermédiaire à fort (Fig.40A-C; phénotypes *lfy* décrits en Introduction). Pour savoir si ces phénotypes résultent d'une modification du comportement de liaison de ces protéines mutées, nous avons produit et analysé les protéines *lfy-4*, *lfy-5*, *lfy-20* et *lfy-28* dans le contexte LFY-C. *lfy-4*, *lfy-5* et *lfy-20*, comme nous pouvions nous y attendre, affichent une diminution d'affinité pour AP1 WT comparativement à la protéine sauvage (Fig.40D). Mais plus étonnant, *lfy-28* (P308L) occasionne lui aussi une forte perte d'affinité alors que comme nous l'avons vu, une autre mutation de ce résidu en alanine (P308A) est sans conséquence (Fig.39A). Cette différence trouve probablement explication dans l'encombrement stérique propre à la leucine et à l'alanine: la leucine est un long résidu qui en prenant la place de la proline créerait une incompatibilité stérique entre l'hélice $\alpha 3$ de LFY-C et l'ADN alors que l'alanine, plus petite, ne perturberait pas cet agencement de l'hélice $\alpha 3$ au niveau de l'ADN.

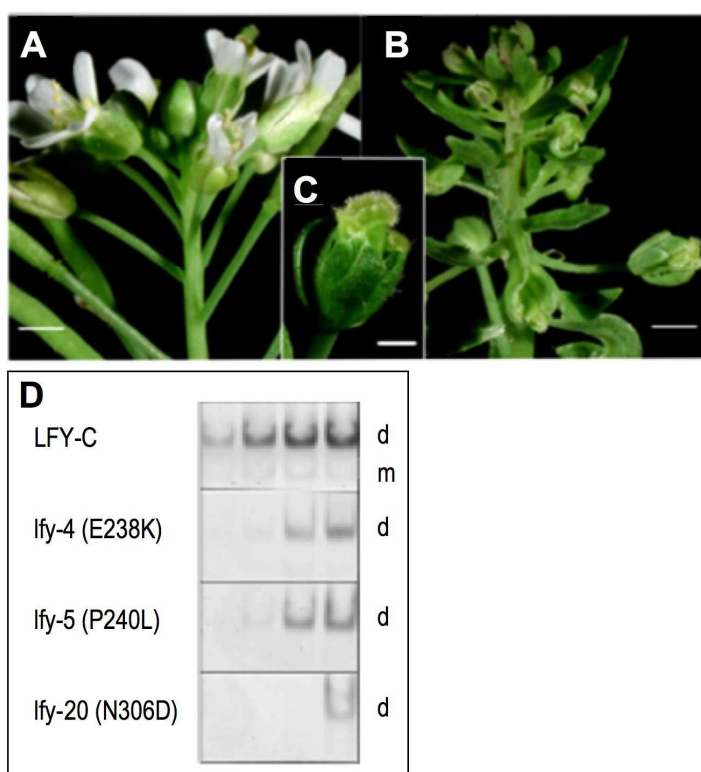


Figure 40: Etude des mutants *lfy* décrits *in planta*

(A-C) Photographies comparatives d'une inflorescence sauvage (A) et mutante *lfy-28* (P308L) (B) d'*Arabidopsis thaliana* et détails d'une fleur *lfy-28* (C); la barre d'échelle correspond à 1mm en (A) et (B) et 0.5mm en (C); (D) Etude de l'effet des mutations *lfy-4*, *lfy-5* et *lfy-20* sur l'affinité de liaison de LFY à AP1. EMSA réalisés avec 10nM AP1 et 100, 250, 750, 2000nM protéines; les positions du dimère (d) et du monomère (m) sont indiquées.

c) Confirmation des bases moléculaires à l'origine de la coopérativité

La structure de LFY-C révèle l'existence d'une interaction monomère-monomère mettant supposément en jeu le résidu Asp280 d'un coté et les résidus His387 et Arg390 de l'autre. Pour confirmer cette hypothèse, nous avons étudié l'effet de mutation dans ces résidus sur le comportement de liaison de LFY-C. La mutation D280K ne semble pas modifier le comportement de liaison de LFY-C à l'ADN puisque la coopérativité et l'affinité de liaison de LFY à l'ADN sont comparables à celles de la protéine sauvage (Fig.41A).

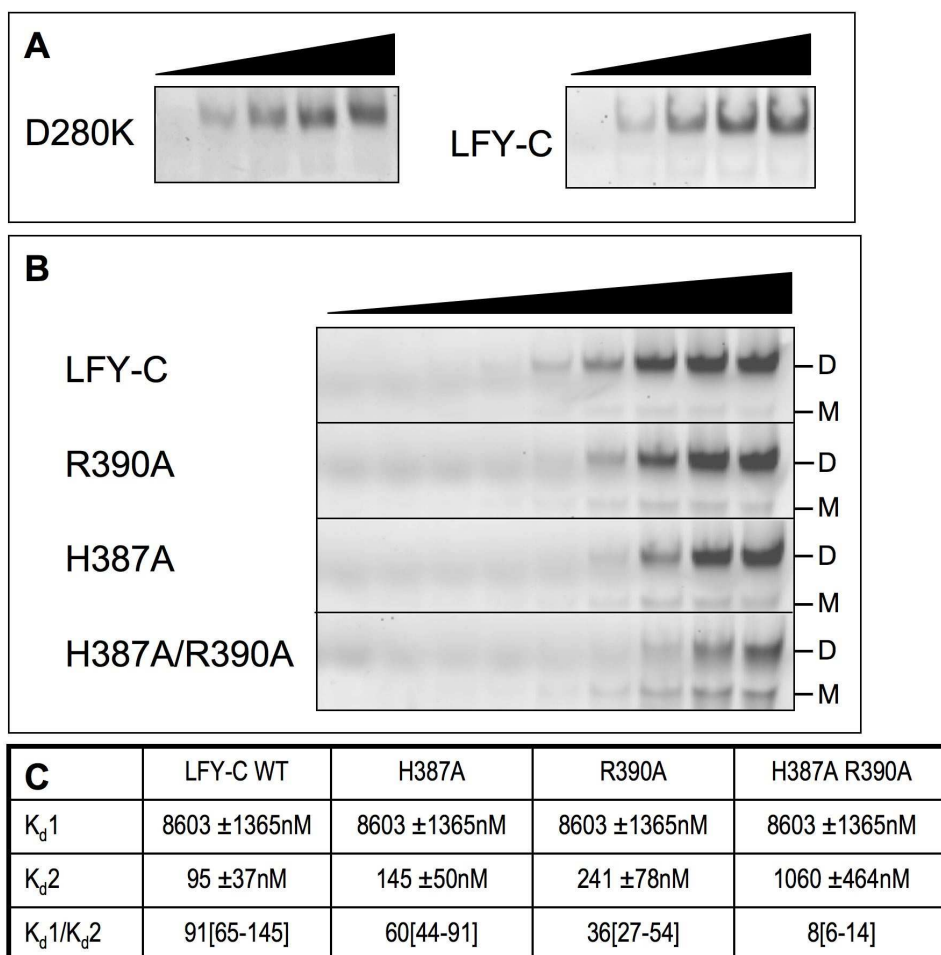


Figure 41: Détermination des résidus impliqués dans la coopérativité de liaison de LFY-C à l'ADN par EMSA

(A) EMSA réalisé avec 10nM AP1 et une gamme croissante de D280K et LFY-C (0, 250, 500, 1000, 5000nM de gauche à droite); (B) EMSA réalisés avec 50nM AP1 et une gamme croissante de protéine H387A, R390A ou H387A/R390A (0, 10, 20, 50, 100, 200, 500, 1000, 2000, 3000nM de gauche à droite). Les positions du dimère (D) et du monomère (M) sont indiquées); (C) Estimation des constantes de dissociation de LFY-C, H387A ou H387A/R390A dans la réaction de liaison de la protéine à l'ADN. Les erreurs indiquent un intervalle de confiance de 95%.

Il ne s'agit donc pas d'une interaction ionique entre les chaînes latérales comme nous l'avions envisagé initialement. En revanche, les simples mutations H387A et R390A ainsi que la double mutation H387A/R390A entraînent une perte de coopérativité, ainsi qu'une perte d'affinité de LFY-C pour l'ADN dans le cas H387A/R390A (Fig.41B). De la même manière qu'il l'avait été décrit pour quantifier la coopérativité de LFY-C, nous avons estimé l'indice de coopérativité de H387A, R390A et H387A/R390A

respectivement à 36, 60 et 8. Ces résultats suggèrent donc que les résidus essentiels à la coopérativité LFY-C/LFY-C sont l'His387 et l'Arg390.

4. Discussion

a) Un nouveau domaine de liaison à l'ADN

Les données structurales confirment que LFY est un facteur de transcription original et spécifique aux plantes. Il présente un nouveau domaine de liaison à l'ADN de type HTH. Jusqu'ici, la grande majorité des gènes homéotiques végétaux codaient des protéines à boîte MADS, alors que de nombreux gènes homéotiques animaux appartiennent à la famille des protéines à motif HTH de type homéodomaine ou domaine paired. Le fait que LFY, le régulateur de ces protéines à boîte MADS, présente des similarités structurales avec les homéoprotéines animales suggère donc que finalement, le même type de DBD ait été adopté par les plantes et les animaux pour orchestrer des transitions développementales majeures.

b) Un domaine extrêmement conservé de la Mousse aux végétaux supérieurs

La structure de LFY-C est très compacte, résultat de nombreuses interactions entre résidus au sein du coeur hydrophobe. La fonctionnalité de LFY-C repose donc sur une organisation tridimensionnelle extrêmement contrainte, ce qui explique certainement une partie de sa grande conservation au cours de l'évolution.

c) Identification des points de contact LFY-C/ADN

La structure a également permis d'identifier les résidus impliqués dans la liaison de LFY à l'ADN. Seuls trois résidus seraient donc impliqués dans une reconnaissance base-spécifique de l'ADN, tandis que douze contactent le squelette phosphate de l'ADN. Mais une considération importante est à prendre en compte ici: le complexe cristallisé se compose d'un oligonucléotide de 20pb et les résidus de LFY-C dont la structure a pu être résolue se limitent au domaine 237-399. Les sept premiers résidus ordonnés à l'extrémité N-terminale du domaine conservé sont donc absents de ce jeu de données. Or, ils pourraient également intervenir dans la liaison de LFY à l'ADN comme le suggère l'étude du comportement de liaison des protéines produites à partir de pCH42 (résidus 241-421) où la délétion de ces résidus (et de quatre résidus supplémentaires) diminue fortement l'affinité de liaison de LFY-C à l'ADN (Fig.23).

L'Arg235 serait une cible particulièrement intéressante à étudier à ces fins car elle est extrêmement bien conservée chez toutes les plantes terrestres et se situe à proximité de l'Arg237, un des trois points de contact base-spécifique déjà établi. D'autre part, en considérant l'oligonucléotide, nous avons vu en chapitre I que LFY-C contacte l'ADN au-delà de 21pb: il est probable qu'à ce niveau également les données obtenues ne soient pas complètes. Notamment, il est intéressant de noter

qu'une prolongation de résidus basiques (Lys253, Lys254) est observée sur une zone externe de LFY-C parallèle au brin d'ADN (Fig.42), laissant supposer que cette zone pourrait être impliquée dans le contact à un plus long fragment d'ADN.

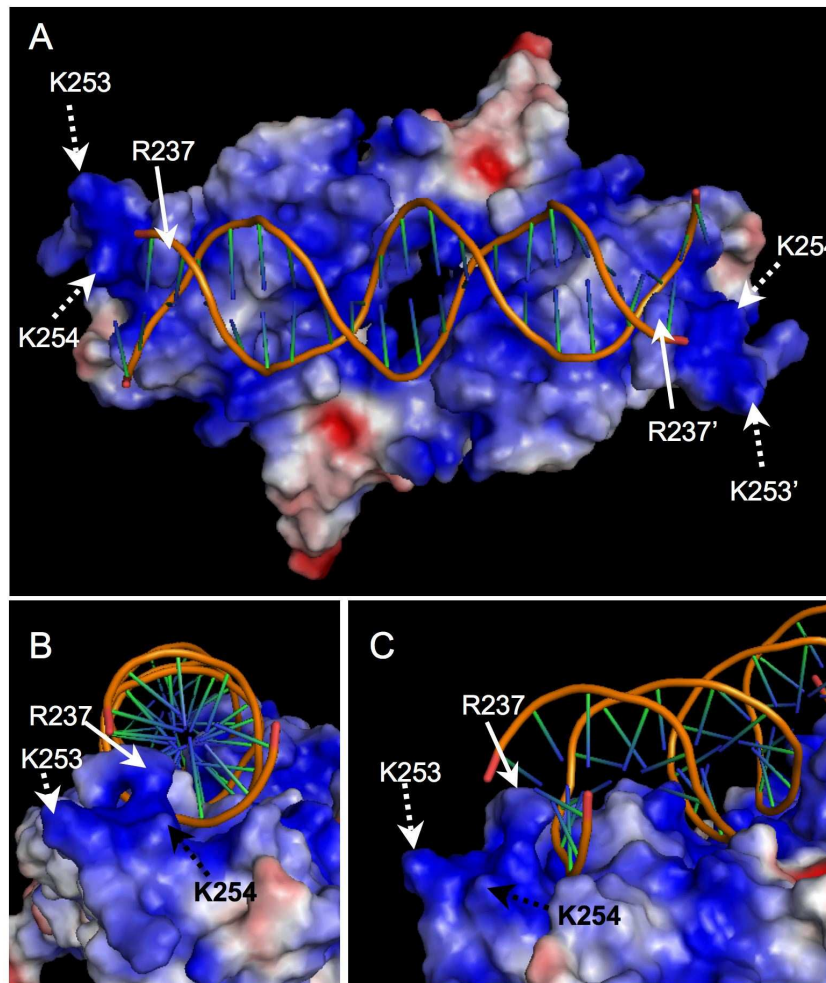


Figure 42: Mise en évidence d'une possible prolongation des points de contact LFY/ADN sur la structure du dimère LFY-C lié à AP1 (travail de F.Parcy)

(A-C) Représentation générale du complexe LFY-C/ADN vu du dessus (A) et gros plans (B et C) sur la zone de prolongement putative de l'ADN mettant en évidence l'insertion de l'Arg237 au niveau du petit sillon de l'ADN (flèche blanche) et la présence de deux lysines (K253 et K254) au niveau de LFY-C (flèche en pointillés blanche). Visualisation à l'aide du logiciel PyMOL (<http://pymol.sourceforge.net/>). Le dimère LFY-C est représenté en surface, les potentiels électrostatiques positifs et négatifs estimés à l'aide du logiciel PyMOL sont représentés respectivement en bleu et en rouge. L'ADN double brin résolu par cristallographie est représenté en orange.

d) Bases moléculaires de la coopérativité

Le mécanisme moléculaire principal à l'origine de la coopérativité a été suggéré par approche cristallographique. Mentionnons que si le cristal a été obtenu avec un oligonucléotide de 20pb, taille diminuant l'affinité de liaison de LFY à l'ADN, des études d'EMSA avec cet oligonucléotide ont montré que LFY-C conservait tout de même un comportement de liaison coopératif (données non montrées). Ce mécanisme reposerait ainsi sur une dimérisation de LFY-C impliquant trois résidus, l'Asp280, l'His387 et l'Arg390. Toutefois, la structure à elle seule ne peut être suffisante pour déterminer la part

jouée par chacun de ces résidus. L'aspartate étant un acide aminé chargé négativement et l'histidine et l'arginine positivement, nous avons initialement pensé que les liaisons mises en jeu ici étaient de nature ioniques. Or, la mutation de l'Asp280 en lysine chargé positivement (D280K) ne semble pas perturber le comportement de liaison de LFY-C (Fig.41A), alors qu'elle devrait avoir un effet répulsif sur l'interaction de nature ionique avec l'His387 et l'Arg390 chargées positivement aussi. Les liaisons sont donc d'une nature autre, révisées comme étant des ponts hydrogène de l'His387 et de l'Arg390 au groupement carbonyle de l'Asp280. L'estimation de l'importance relative de chacun de ces résidus sur la coopérativité a donc nécessité une étude biochimique complémentaire. La double mutation H387A/R390A affiche l'effet le plus fort, soit une perte de coopérativité d'un facteur 11 environ comparativement à la protéine LFY-C sauvage, alors que la simple mutation D280K n'a aucun effet. Les simples mutations H387A et R390A n'ont quant à elles qu'un léger effet sur l'affinité de liaison et la coopérativité, signifiant probablement que chaque résidu joue un rôle équivalent, et que l'absence de l'un est en large partie compensée par l'autre. L'His387 et l'Arg390 sont ainsi les deux déterminants majeurs de la coopérativité dans le contexte de LFY-C cristallisé (résidus 237-399).

Soulignons que la coopérativité résiduelle chez le double mutant H387A/R390A (indice de coopérativité égal à 8) suggère l'intervention d'un autre mécanisme. Il pourrait s'agir comme nous l'avons évoqué en discussion du premier chapitre d'une implication de l'ADN. La liaison du premier monomère (préférentiellement sur le demi-palindrome 3') pourrait occasionner un léger changement de conformation de l'ADN qui pourrait mettre en jeu la courbure de l'ADN observée par la résolution de structure du complexe LFY-C/ADN. Alternativement, il pourrait également s'agir d'interactions plus faibles entre monomère qui n'ont pas été identifiées, du type interactions de Van der Waals réparties le long de la surface. La structure révèle en effet que les deux monomères sont relativement bien imbriqués l'un dans l'autre.

e) Par quelle voie LFY-C lie-t-elle l'ADN?

Les données structurales, montrant que LFY-C établit des points de contact entre chaque monomère, soulèvent la question de la voie empruntée par LFY-C pour lier l'ADN. En effet, un facteur de transcription capable de se dimériser peut lier l'ADN selon deux voies distinctes: la voie dite 'monomérique' se définit par une liaison indépendante de chacun des deux monomères à l'ADN (puis dimérisation sur l'ADN), alors que la voie 'dimérique' signifie que l'ADN est lié directement par le dimère (Fig. 43; Kohler et al., 1999). Même si ces deux voies ayant la même finalité (2 protéines liées à l'ADN) présentent un même bilan thermodynamique (Cranz et al., 2004), la voie monomérique présente un avantage certain par rapport à la voie dimérique.

La voie monomérique augmente en effet la rapidité de prospection de l'ADN. Considérant que la reconnaissance de l'ADN se fait essentiellement par mouvement de diffusion brownienne (von Hippel and Berg, 1989), le monomère, développant moins d'interactions électrostatiques non spécifiques avec

l'ADN que le dimère, est moins interrompu dans son déplacement jusqu'à sa cible d'ADN. Selon un même raisonnement, il est moins sujet à des séquestrations au sein de complexes protéiques non spécifiques (Cranz et al., 2004; Kohler et al., 1999). De plus, la voie monomérique laisse l'opportunité au premier monomère d'interagir avec d'autres partenaires protéiques sur l'ADN, hétérodimères qui augmentent potentiellement le nombre de séquences d'ADN reconnues (Kohler and Schepartz, 2001; Vinson et al., 2002).

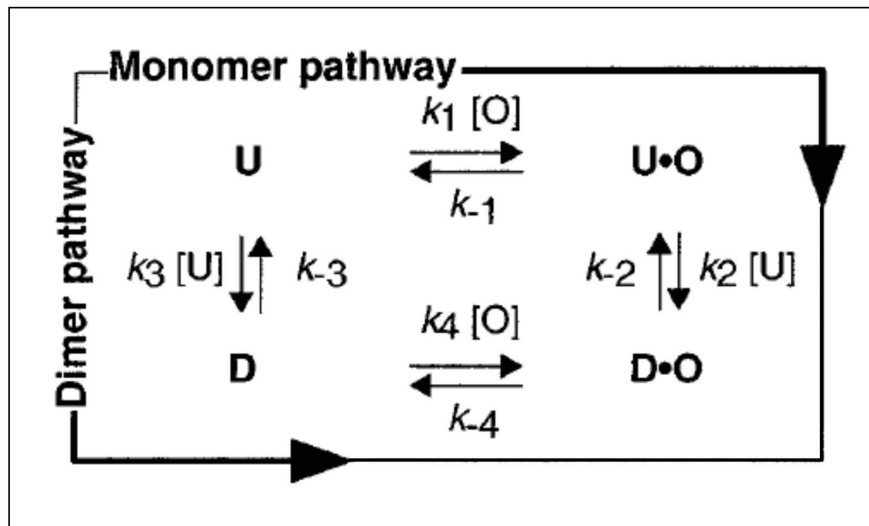


Figure 43: Illustration des deux voies de liaison de l'ADN pouvant être empruntées par une protéine dimérique (d'après Kohler et al., 1999)

U = Monomère; D = dimère; O = oligonucléotide d'ADN.

Par déduction, la voie dimérique est moins avantageuse que la voie monomérique, augmentant la probabilité de liaisons non spécifiques avec l'ADN ou des partenaires protéiques. Toutefois, la voie dimérique peut présenter un intérêt chez certaines protéines comme de nombreuses bZIP, où la dimérisation stabilise les deux monomères initialement non structurés. Cet intérêt est néanmoins limité dans la mesure où l'absence de repliement augmente le rayon de capture de la protéine et par conséquent sa capacité à contacter l'ADN (Cranz et al., 2004; Kohler et al., 1999). De plus, la voie dimérique peut être bénéfique si elle empêche la formation d'hétérodimères sur l'ADN avec des partenaires protéiques jouant un effet négatif sur l'activité de la protéine considérée.

Dans le contexte d'une cellule, l'excès de protéines et d'ADN favorise la formation de complexes non spécifiques. Ces complexes d'une durée de vie relativement longue diminuent d'autant plus l'efficacité de la voie dimérique et augmentent la possibilité d'une réponse transcriptionnelle inappropriée (Kohler et al., 1999). Ces éléments expliquent probablement pourquoi de nombreux facteurs de transcription tels le répresseur LexA, ATF-2, Max, Jun et Fos empruntent la voie monomérique (Kohler and Schepartz, 2001).

Diverses stratégies ont été utilisées afin d'identifier la voie sollicitée par différents facteurs de transcription. Kohler et Schepartz (2001) par exemple ont étudié l'efficacité de FRET (*Fluorescence*

Resonance Energy Transfer) d'un mélange Jun-fluorescein et Fos-rhodamine en présence ou non d'ADN. Mentionnons que Jun et Fos sont deux protéines bZIP qui lient essentiellement l'ADN sous forme d'hétérodimères, bien que les homodimères puissent également exister. Le taux d'hétérodimérisation obtenu par FRET, plus élevé en présence d'ADN, suggère que ces deux protéines empruntent la voie monomérique. Une autre stratégie consiste à découpler les différents acteurs de ces voies en étudiant par exemple l'effet d'une dimérisation obligatoire de la protéine en solution par l'établissement de ponts disulfures entre chaque monomère, ou à l'inverse l'effet d'une mutation dans les résidus impliqués dans la dimérisation de la protéine (Cranz et al., 2004). Enfin, il est possible par une étude plus fine de déterminer les différentes constantes cinétiques mises en jeu pour en déduire la voie empruntée (Fig.43).

Dans notre cas, le fait que LFY-C soit capable de lier l'ADN sous forme monomérique (complexe C1) suggère que LFY-C emprunte la voie monomérique, hypothèse soutenue par les résultats de SEC analytique montrant que LFY-C seule en solution est principalement monomérique. Toutefois, nous ne pouvons pas exclure que LFY-C ne puisse également contacter l'ADN selon la voie dimérique, même si cette voie serait largement secondaire. En effet, en SEC analytique, j'ai parfois constaté un très léger pic situé au niveau du volume d'élution correspondant à la masse théorique du dimère (Fig.19D flèche en pointillés).

f) Un complexe d'ordre supérieur ?

LFY-C à très forte concentration occasionne la formation d'un complexe d'ordre supérieur, formation qui serait facilitée par la mutation P308A. Quelle est la nature de ce complexe ? Sur quels contacts repose-t-il ? Il pourrait s'agir d'une agrégation forcée de la protéine due à sa trop forte concentration (ou à une instabilité de repliement dans le cas de P308A), ou d'un complexe ayant une vraie réalité biologique, mettant en jeu des interactions protéine-protéine et/ou protéine/ADN. Ce point sera abordé en discussion générale.

III. Caractérisation fonctionnelle de la protéine LFY FL

1. Introduction

Nous avons établi que LFY-C était majoritairement monomérique en solution et qu'elle se dimérisait selon un mécanisme coopératif sur l'ADN. Les données structurales ont suggéré que deux résidus seraient les principaux acteurs de cette coopérativité, ce qui a été validé par approche biochimique. Conscients des limites de travailler sur un sous-domaine de la protéine, nous avons voulu savoir si cette coopérativité est une réalité dans le contexte LFY FL, et si tel est le cas, si elle intervient *in planta* dans la transition florale comme nous l'avons postulé dans le précédent chapitre. Est ce que le comportement de liaison de LFY FL peut se résumer au comportement de liaison de LFY-C, ou est ce que d'autres résidus exclus de ce domaine tels le domaine conservé N-terminal interviennent également dans la dimérisation et/ou la liaison à l'ADN ?

2. Etude *in vitro* du comportement de liaison à l'ADN de LFY FL

a) Une protéine difficile à produire

De nombreux essais ont été entrepris dans le but de produire et purifier la protéine LFY FL. Divers facteurs ont été testés en ne conservant qu'une constante majeure, le système de production hétérologue en bactéries. Les natures du vecteur, de la souche, du milieu de culture, du tampon de purification, la température et le temps d'induction et la concentration en inducteur IPTG ont quant à eux été modulés afin d'obtenir le meilleur rendement possible en LFY FL. Le tableau 4 récapitule ces essais encore très insatisfaisants.

Une remarque particulière est à faire ici: au cours de mon DEA, je suis parvenue à obtenir une protéine LFY FL en très faible quantité mais dont il était possible de cliver l'étiquette solubilisatrice GST. Une partie des résultats présentés ci-dessous ont été obtenus à cette période. Au cours de ma première année de thèse, j'ai mis beaucoup d'énergie à tenter de réitérer et d'optimiser ces productions mais sans succès. C'est pour cette raison que nous avons pris la décision d'étudier qu'un sous-domaine de la protéine (le domaine LFY-C). Le travail de production de la protéine pleine longueur n'a été repris que très récemment et je m'attelle tout particulièrement à obtenir une protéine sans étiquette GST (qui peut induire une dimérisation artificielle de la protéine d'intérêt). La présence de cette étiquette aide beaucoup à solubiliser la protéine LFY mais il est très difficile de s'en débarrasser par la suite, son clivage déstabilisant fortement LFY FL et conduisant à son agrégation. Ma stratégie à ces fins est double: essayer de produire une protéine LFY-GST en grande quantité pour améliorer le rendement en protéines LFY après clivage de l'étiquette GST (travail à partir du vecteur pCH52, tab.4), ou parvenir à induire l'expression d'une protéine LFY sans étiquette GST (travail à partir du vecteur pCH32, tab.4).

Tableau 4: Récapitulatif des essais de production et purification de la protéine LFY FL
6his= étiquette 6histidines, trx = étiquette *thioredoxin carrier*, GST= glutathione S-transferase,
TEV = site de clivage à la protéase TEV, thrombine = site de clivage à la thrombine.

Nom du vecteur	Vecteur d'origine	Fusion traductionnelle (étiquette(s) de solubilisation et/ou de purification)	Poids moléculaire (en kDa)	Expression
pCH2	pGEX-4T1 +étiquette 6his C-terminale	GST-thrombine-LFY-6his	Avec GST: 74 Après clivage GST: 45	Induction satisfaisante, protéine à 0,75mg/mL avant clivage GST mais problème d'agrégation après clivage.
pCH8	pGEX-4T1 +étiquette 6his C-terminale	GST-thrombine-GFP-LFY-6his	Avec GST: 104 Après clivage GST: 75	Pas d'induction
pJUL1	pGEX-4T1 +étiquette 6his C-terminale	GST-thrombine-LFY-GFP-6his (ADNc de la GFP inséré dans l'ADNc de LFY)	Avec GST: 102 Après clivage GST: 73	Induction satisfaisante, protéine à 0,75mg/mL avant clivage GST mais difficulté à séparer les protéines clivées des protéines non clivées.
pCH15	pET21b +étiquette 6his N-terminale	6his-LFY-6his	56	Pas d'induction
pCH17	pET21b +étiquette 6his N-terminale	6his-LFYtronquée-6his (délétion de la région hydrophobe C-terminale)	55	Pas d'induction
pETH18	pETM-30	6his-GST-TEV-LFY-6his	Avec GST: 77 Après clivage GST: 48	Induction satisfaisante, protéine à 1,2mg/mL avant clivage GST mais problème de pureté et problème d'agrégation après clivage.
pCH52	pETM-30	6his-GST-TEV-LFY	Avec GST: 76 Après clivage GST: 47	Belle induction, travail de purification en cours d'optimisation.
pCH56	pETtrx1a	6his-trx-TEV-LFY	Avec GST: 61 Après clivage GST: 48	Pas d'induction
pCH57	pETtrx1b	trx-6his-TEV-LFY	Avec GST: 61 Après clivage GST: 48	Pas d'induction
pCH32	pETM-11	6his-LFY-6his	51	Induction en cours d'optimisation

b) Premières caractérisations fonctionnelles

1- Fonctionnalité et caractérisation quantitative de la liaison de LFY FL à l'ADN

De la même manière que pour la protéine LFY-C, la fonctionnalité de la protéine LFY FL a été validée par AF et EMSA (Fig.44). Le K_d de la réaction de liaison de LFY FL à AP1 déterminé par ces deux techniques atteint une valeur de l'ordre de 100nM également, tandis que le K_d de LFY FL pour AG-11 est estimé à 200nM par EMSA. Ces données traduisent une affinité et une spécificité de liaison à l'ADN similaire entre LFY FL et LFY-C.

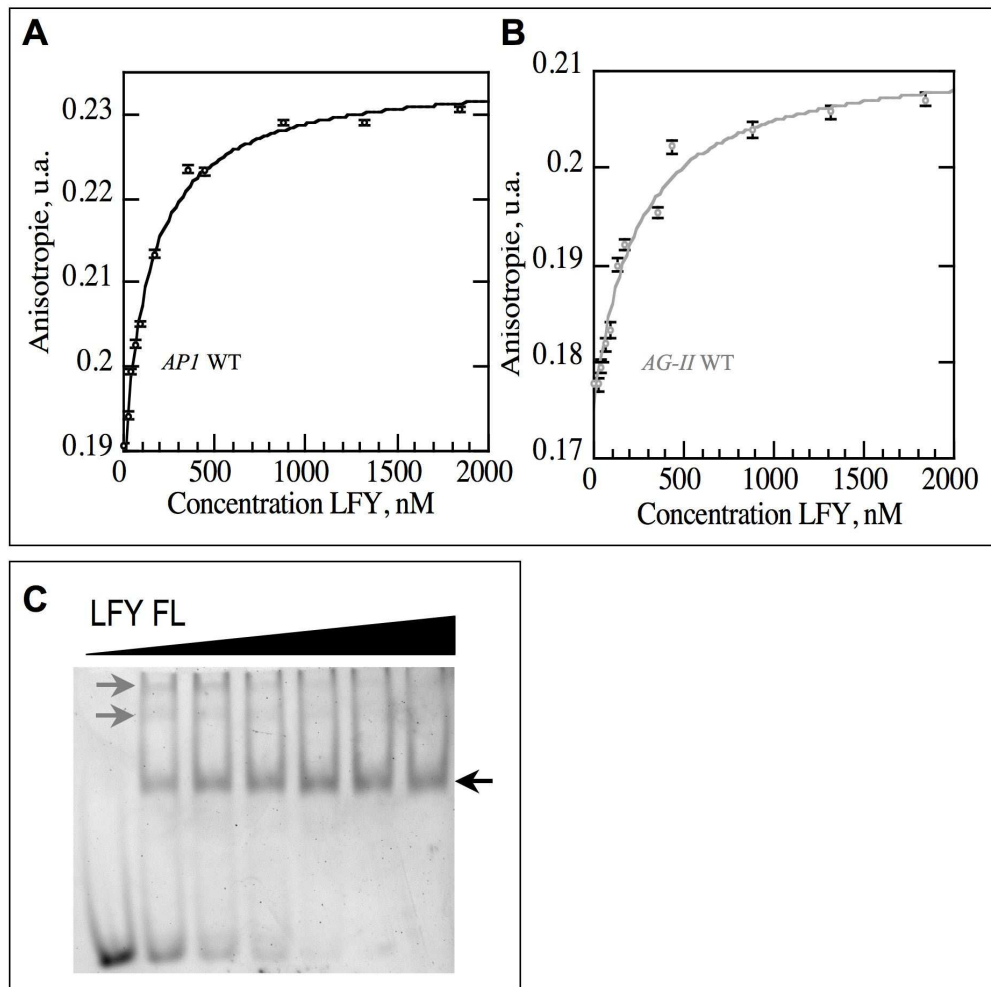


Figure 44: Test de fonctionnalité de LFY FL par EMSA et AF

(A-B) Courbes de saturation de LFY à *AP1* (A; en noir) et à *AG-II* (B; en gris) obtenues par AF. Les courbes représentent l'ajustement des points expérimentaux à l'équation théorique. Les barres d'erreur correspondent aux erreurs standards (N=60); (C) EMSA réalisé avec 10nM *AP1* et une concentration croissante en LFY-GST (0, 30, 80, 130, 270, 530, 1300nM de gauche à droite). La flèche noire signale le complexe majoritaire, les flèches grises les complexes d'ordre supérieur.

2- Cinétique d'interaction LFY/ADN

Un des avantages de la technique d'AF est de permettre de suivre une cinétique d'association ou de dissociation en temps réel. Pour estimer le temps mis par LFY à se lier à l'oligonucléotide *AP1* WT, nous avons suivi l'évolution de l'AF de l'oligonucléotide suite à l'injection d'un excès de LFY (Fig.45A). En moins de 30sec, il y a association puisque l'AF atteint un plateau. Ces données justifient donc que les mesures d'AF soient prises peu de temps après le mélange des deux réactifs.

3-Estimation du k_{-} de l'interaction LFY/oligonucléotide *AP1* WT

La technique d'AF nous a également donné accès au k_{-} (constante de vitesse de dissociation) de la réaction. Pour cela, nous avons effectué une expérience de chasse en injectant un excès (50X) d'oligonucléotide *AP1* WT non fluorescent dans une solution où tous les oligonucléotides *AP1* fluorescents sont liés à LFY. Là encore, l'équilibre est atteint au bout de quelques secondes (Fig.45B).

La confrontation des équations théoriques (Wachstock and Pollard, 1994) aux points expérimentaux permet d'estimer la valeur du k^- à $0,01 \text{ sec}^{-1}$ pour deux préparations de LFY indépendantes.

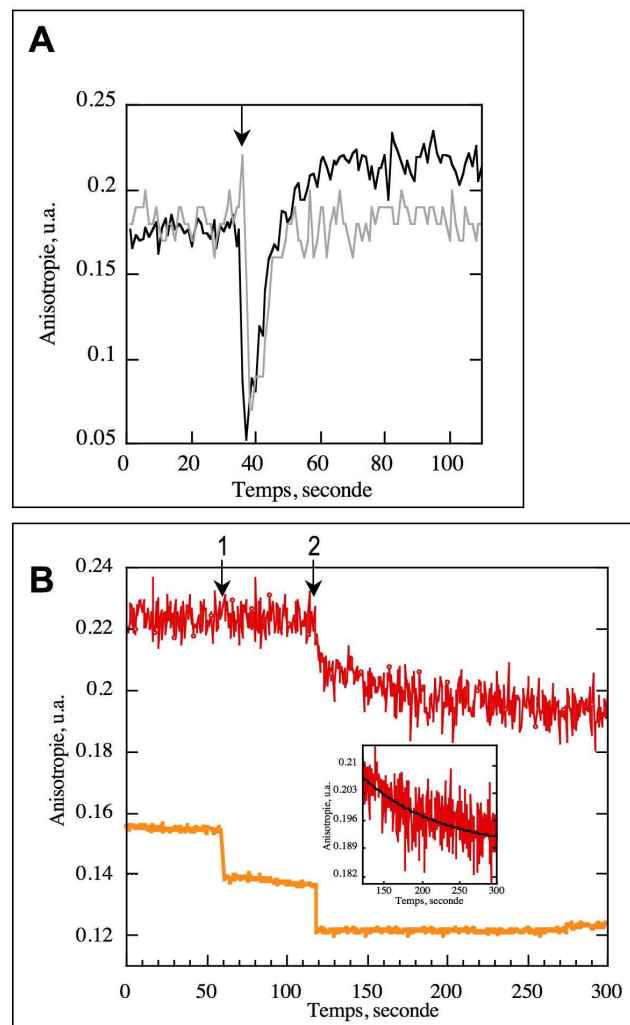


Figure 45: Etude des cinétiques de liaison et de dissociation de LFY FL à AP1 par AF

(A) Etude de la cinétique de liaison de LFY à AP1. La flèche signale l'injection de LFY (courbe noire) ou de tampon contrôle (courbe grise). Après injection de LFY, et une fois la perturbation liée à l'injection passée, l'AF atteint un plateau de valeur égale à celle obtenue pour la même concentration finale en LFY dans l'expérience de saturation (figure 44A, AF à 450 nM en LFY). L'injection de tampon ne modifie que transitoirement l'AF; (B) Détermination du k^- de l'interaction LFY/AP1. La flèche 1 signale l'injection de tampon contrôle, la flèche 2 l'injection d'un excès 50 fois d'oligonucléotides AP1 non marqués par rapport aux oligonucléotides marqués. En rouge, valeurs d'AF, en orange, valeurs de fluorescence. L'agrandissement représente l'ajustement monoexponentiel réalisé sur les points expérimentaux après injection d'oligonucléotides non marqués. Les injections diluent les oligonucléotides marqués, induisant une chute de la fluorescence. Les mesures sont moyennées à partir de trois expériences indépendantes.

4-Mise en évidence d'une coopérativité de liaison de LFY FL sur l'ADN

L'EMSA reporté en figure 44C illustre la formation de différents complexes LFY FL/AP1. Pour tester si le complexe majoritaire peut correspondre au dimère de LFY FL lié à AP1, nous avons étudié le comportement de liaison de LFY FL à AP1m1, AP1m2 et AP1m3, oligonucléotides qui occasionnaient une accumulation du complexe monomérique de LFY-C sur AP1 en perturbant la formation du dimère (Fig.28C). L'EMSA ainsi réalisé révèle qu'effectivement, la liaison de LFY FL à ces

oligonucléotides mutés entraîne l'accumulation d'un plus petit complexe difficilement visible à partir du mélange LFY FL/AP1 WT (Fig.46). Ces résultats suggèrent fortement que le complexe majoritaire formé par LFY FL et AP1 soit un dimère et que LFY FL puisse lier l'ADN sous forme monomérique. Notons cependant qu'il n'est pas totalement exclu qu'il s'agisse de complexes d'ordre supérieur comme par exemple deux protéines LFY FL dans le petit complexe et trois ou quatre protéines LFY FL dans le complexe majoritaire. Quoiqu'il en soit, si notre hypothèse se révèle exacte, le fait que le petit complexe soit toujours largement minoritaire devant le complexe plus important implique que la coopérativité existe également dans le contexte LFY FL.

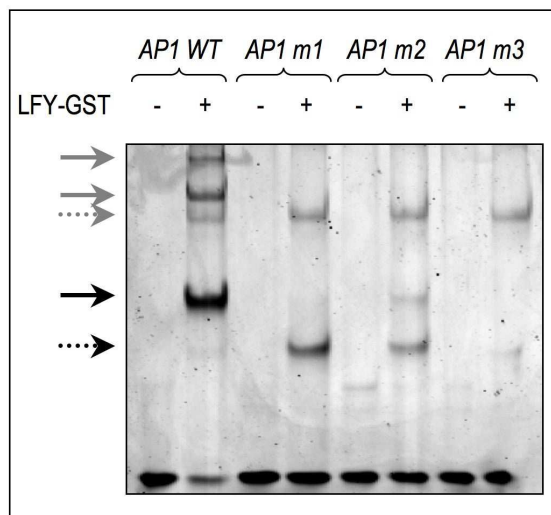


Figure 46: Etude de la coopérativité de liaison de LFY FL à AP1 par EMSA

EMSA réalisé avec 10nM d'oligonucléotide AP1 WT, m1, m2 ou m3, en présence (+) ou non (-) de 1000nM LFY-GFP-GST. La flèche continue noire signale le dimère putatif LFY lié à AP1, la flèche en pointillés noirs signale le monomère putatif, les flèches grises signalent les complexes de masses apparentes plus élevées, continues dans le cas des complexes potentiellement spécifiques d'AP1, et en pointillés dans le cas du complexe potentiellement non spécifique.

5-Mise en évidence de complexes d'ordre supérieur à faible concentration en LFY FL sur AP1

Les EMSA réalisés avec LFY FL laissent apparaître des complexes de taille apparente plus forte sur AP1 en plus des monomère et dimère putatifs (Fig.44C et 46, flèches grises). Si l'EMSA reporté en figure 44C révèle la présence de deux complexes supérieurs, l'EMSA reporté en figure 46 en révèle trois (complexe supplémentaire signalé par une flèche grise en pointillés). Cette différence pourrait provenir de la qualité de la protéine qui a occasionné ce retard supplémentaire après un séjour d'une nuit à 4°C. Le fait que contrairement aux autres complexes d'ordre supérieur, ce complexe soit observé quelque soit la nature de l'oligonucléotide (AP1 WT, m1, m2 ou m3) suggérerait qu'il n'est pas spécifique de la séquence d'AP1. La découverte de ces complexes ouvre de nouvelles perspectives d'étude que je discuterai en discussion générale.

6-Mise en évidence de sites de liaison putatifs de LFY au niveau du promoteur d'*AP1*

A partir de la séquence consensus CCANTG de liaison de LFY à l'ADN, nous avons identifié deux séquences putatives de liaison de LFY au promoteur d'*AP1* supplémentaires. La séquence identifiée par Parcy et collaborateurs (1998), que je dénommais jusqu'ici '*AP1*' correspond au site '*AP1 bs1*'. Ces deux sites putatifs seront décrits en tant qu'*AP1bs2*' et '*AP1bs3*'.

Dans le but de comprendre l'éventuelle importance biologique de ces deux sites supplémentaires, j'ai entrepris une caractérisation biochimique de liaison de LFY FL à ces différents sites. La figure 47 révèle que LFY FL est capable de lier *AP1bs2* et *AP1bs3* mais que l'affinité de ces interactions est nettement moins forte que celle entre LFY FL et *AP1bs1*: les K_d de LFY FL sont estimés à plus de $2\mu\text{M}$ pour *AP1bs2* et à environ $1\mu\text{M}$ pour *AP1bs3*.

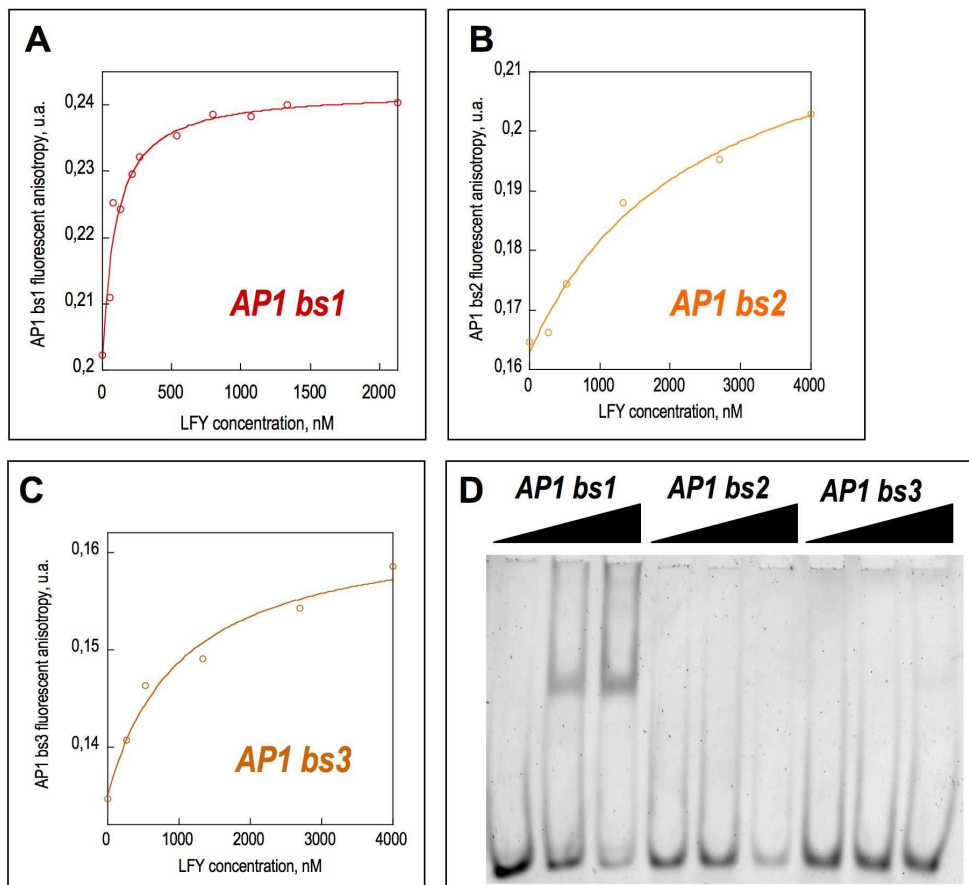


Figure 47: Caractérisation de la liaison de LFY FL au promoteur d'*AP1* par EMSA et AF
 (A-C) Courbes de saturation de LFY à *AP1 bs1* (A), à *AP1 bs2* (B), et à *AP1 bs3* (C) obtenues par AF. Les courbes représentent l'ajustement des points expérimentaux à l'équation théorique. La faible concentration en LFY FL n'a pas permis d'atteindre le plateau de saturation avec *AP1bs2* et *AP1bs3*;
 (D) EMSA réalisé avec 10nM *AP1 bs1*, *bs2* et *bs3* (précisé) et une concentration croissante en LFY FL (0, 30, 80nM de gauche à droite).

3. Etude de l'homodimérisation de LFY FL *in vivo* chez la levure

Travailler *in vitro* permet de simplifier le modèle d'étude aux seules molécules en présence et de contrôler tous les paramètres extérieurs tels que le tampon, la concentration saline ou encore la température. Cette approche apporte donc des évidences directes. Cependant une telle simplification peut facilement être remise en cause. Aujourd'hui en effet, de nombreuses données sur l'organisation de l'ADN dans son environnement nucléaire montrent que cet environnement peut considérablement influencer sur la capacité des facteurs de transcription à accéder à leur cible d'ADN (Gregory et al., 1998; Gregory et al., 2001). On pourrait également concevoir que ces interactions puissent être court-circuitées par d'autres acteurs protéiques présents dans l'environnement cellulaire qui pourraient séquestrer la protéine d'intérêt, entraîner sa dégradation, ou encore entrer en compétition avec elle pour des mêmes séquences cible d'ADN. Enfin, la question de la concentration de LFY *in cellulo* est également d'importance majeure: en fonction de l'affinité de liaison de LFY à ses cibles, des interactions que nous observons *in vitro* peuvent n'avoir aucune raison d'être *in vivo*.

Pour toutes ces raisons, il est impératif de valider *in vivo* toute donnée obtenue *in vitro*. Afin de valider la pertinence de la dimérisation de LFY FL *in vivo*, nous avons donc entrepris une première série d'expériences à partir d'un organisme simple, de croissance rapide, et de manipulation génétique aisée, la levure.

a) Mise en évidence d'une dimérisation de LFY sur l'ADN par technique de double hybride

Pour tester la dimérisation de deux protéines chez la levure, il existe une technique 'double hybride' basée sur la fusion de l'une de ces protéines à un domaine de liaison à l'ADN (GAL4-DBD qui lie les UAS, *Upstream Activator Sequence*, GAL4), et de l'autre à un domaine activateur de la transcription (GAL-AD par exemple). Si les deux protéines interagissent, le rapprochement des domaines de liaison à l'ADN et d'activation de la transcription conduit à l'activation d'un gène rapporteur dont l'activité peut être quantifiée. On utilise en général pour cela le gène *lacZ* placé sous le contrôle des UAS GAL4. Il existe également une technique dite de simple hybride applicable à un facteur de transcription (TF) où le facteur est toujours fusionné à un domaine activateur de la transcription mais où le gène rapporteur est placé cette fois sous le contrôle du site de liaison du facteur.

Dans un premier temps, nous avons essayé de mettre en évidence l'homodimérisation de LFY par la technique du double hybride mais aucune interaction n'a pu être observée.

Les données de biochimie suggérant que la dimérisation de LFY ait lieu seulement en présence d'ADN, nous avons utilisé une technique intermédiaire entre simple et double hybride. Pour cela, nous avons eu recours à un gène rapporteur placé sous le contrôle des séquences régulatrices d'*AP1* contenant les sites *bs1* et *bs2*. On savait que ce promoteur pouvait être transactivé par LFY fusionné au domaine activateur constitutif VP16 (LFY-VP16) mais sans savoir quel était l'état d'oligomérisation de

LFY au cours de cette activation (Parcy et al., 1998). Pour avoir accès à cette information, nous avons utilisé deux protéines LFY individuellement inactives: d'une part une protéine LFY sauvage mais non fusionnée au domaine VP16 et d'autre part une protéine *lfy*-VP16 mutée dans son domaine C-terminal, mutation qui compromet sa liaison à l'ADN.

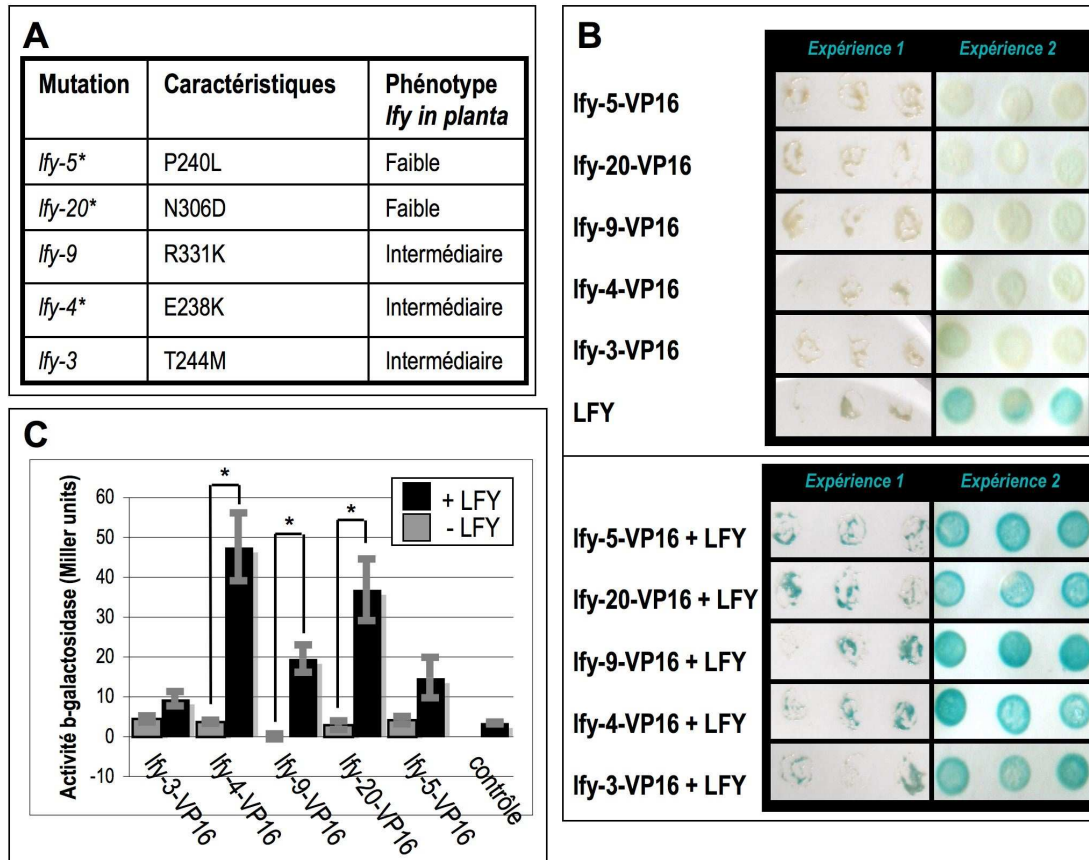


Figure 48: Mise en évidence d'une homodimérisation de LFY chez la levure

(A) Tableau récapitulatif des différents mutants *lfy* utilisés pour l'expérience en levure. Le mode de liaison des mutants signalés d'une astérisque a déjà été évalué dans le contexte LFY-C par EMSA (Fig. 40D); (B) Etude qualitative par simple hybride de la capacité de différentes protéines LFY sauvage et mutées (précisées) à induire l'expression du gène rapporteur *LacZ* placé sous le contrôle du promoteur d'*AP1*. Les résultats de deux expériences indépendantes sont représentés. L'activité est visualisée par suivie de la coloration bleue. Les expériences contrôles sont représentées en haut, les mélanges de *lfy*-VP16 et LFY sont représentés en bas; (C) Etude quantitative par simple hybride des constructions testées en (B). Le contrôle est obtenu à partir de l'expression du vecteur vide. Les barres d'erreur représentent les erreurs standard à la moyenne (SEM) N=3. Les astérisques signalent les écarts significatifs ($p < 0,05$). Les activations résultantes de la co-expression de LFY et *lfy*-VP16 sont toutes significativement plus fortes que celle observée avec LFY exprimé seul.

Plusieurs allèles mutants *lfy* ont ainsi été testés en parallèle, mutants dont la gravité des phénotypes est connue *in planta*, et dont les données structurales ont permis de comprendre en partie l'origine de la déficience (Fig.48A). Les mutations *lfy-5*, *lfy-20* et *lfy-4* touchent des résidus situés à proximité de l'Arg237 ou de la Lys307, toutes deux impliquées dans des contacts base-spécifique au niveau de l'ADN. Les mutations *lfy-3* et *lfy-9* affectent quant à elles les résidus Thr244 et Arg331 qui pourraient être impliqués dans le bon repliement de domaine C-terminal, et qui, remarquablement, interagissent l'un avec l'autre dans le contexte sauvage.

L'expression des constructions effectrices LFY ou lfy-VP16 seules n'active pas l'expression du gène rapporteur *AP1::LacZ*. Une exception est tout de même observée à partir de l'expression de la construction LFY seule dans l'expérience 2. Un temps d'incubation prolongé en présence du substrat de l'activité portée par le gène *LacZ* (activité β -galactosidase) a conduit à l'observation d'une induction de l'expression d'*AP1* à partir de LFY seul. Néanmoins, cette induction est probablement très faible étant donné qu'elle n'est pas observée dans l'expérience 1, ni dans une expérience similaire décrite dans Parcy et collaborateurs (1998). En revanche, la co-expression de LFY et lfy-VP16 occasionne une activation forte du gène rapporteur qui à partir de trois allèles lfy-VP16 différents est significativement différente de celle observée en absence de l'une ou l'autre des deux constructions effectrices (Fig.48B et C). La co-expression entraîne donc une synergie de réponse de LFY et lfy-VP16, résultat cohérent avec une dimérisation possible de LFY/lfy-VP16 dépendante de l'ADN.

b) Mise en évidence d'une implication des résidus His387 et Arg390 dans la capacité de LFY FL à activer AG par simple hybride (travail d'E. Thévenon)

L'étude du domaine LFY-C a montré l'importance des résidus His387 et Arg390 sur sa capacité à se dimériser sur l'ADN. De façon à tester si cette propriété est importante dans le contexte de la protéine entière et dans le noyau d'une cellule vivante, nous avons testé si des mutations dans ces résidus modifiaient la capacité de LFY à activer un gène cible dans la levure. Pour cela, Emmanuel Thévenon, technicien dans notre équipe, a testé les capacités du mutant simple H387A ou du mutant double H387A/R390A fusionnés au domaine activateur VP16 à activer une construction rapportrice *AG::LacZ* chez la levure. En comparant les résultats à ceux obtenus à partir d'un contrôle positif avec la protéine LFY sauvage, nous constatons que ces deux mutants occasionnent une baisse de la capacité de LFY à activer AG, baisse statistiquement significative dans le cas du double mutant (Fig.49). Ces données suggèrent donc fortement que la dimérisation de LFY via l'His387 et l'Arg390 soit importante à l'activité de LFY. L'étude du simple mutant Arg390 sera réalisée dans ces mêmes conditions prochainement.

4. Etude de l'homodimérisation de LFY FL *in planta*

Si les données dans la levure constituent déjà un apport *in vivo*, l'étape ultime de notre travail est d'étudier l'importance de LFY dans son contexte naturel.

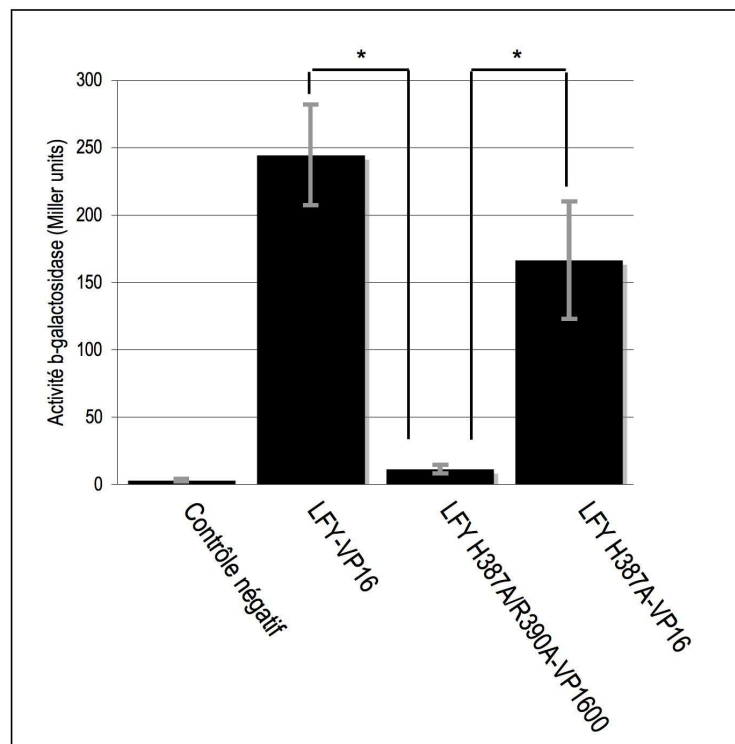


Figure 49: Mise en évidence d'une implication des résidus His387 et Arg390 dans la capacité de LFY FL à activer AG par technique de simple hybride (travail d'E. Thévenon)

Etude quantitative de la capacité de versions sauvage ou mutées de LFY fusionnées au domaine activateur VP16 à induire l'expression du gène rapporteur *LacZ* placé sous le contrôle du promoteur d'AG. Le contrôle est obtenu à partir de l'expression du vecteur vide. Les barres d'erreur représentent les erreurs standard de la moyenne (SEM) N=10. Les astérisques signalent les écarts significatifs ($p < 0,05$).

a) Mise en évidence d'une dimérisation de LFY chez *A. thaliana*

Pour tester si LFY est capable de s'homodimériser dans son contexte natif, nous avons reproduit l'équivalent de l'expérience de simple hybride réalisée avec LFY et *lfy*-VP16 chez *A. thaliana*, à la différence qu'ici les constructions sont placées sous le contrôle du promoteur natif de LFY. Pour cela, j'ai réalisé différents croisements décrits en matériel et méthodes à partir de plantes F0 hémizygotes *lfy*-5-VP16/., *lfy*-12/*lfy*-12 accession *Columbia* (gracieusement fournies par Detlef Weigel, Institut Max Planck, Tuebingen, Allemagne) pour obtenir des plantes de génotype *lfy*-5-VP16/*lfy*-5-VP16, *lfy*-12/*lfy*-12 et des plantes de génotype *lfy*-5-VP16/*lfy*-5-VP16, LFY/LFY.

Avant de m'intéresser aux phénotypes de ces plantes, je vais décrire les phénotypes des plantes mutantes *lfy*-5 et des plantes LFY-VP16 nécessaires à l'interprétation de cette expérience.

La mutation *lfy*-5 obtenue par criblage EMS a été décrite dans l'article de Weigel et collaborateurs (1992). Cette mutation occasionne un phénotype *lfy* faible chez *A. thaliana*. Les fleurs les plus touchées ressemblent à des structures pouvant être intermédiaires entre une inflorescence et une fleur, avec des organes parfois surnuméraires de type feuilles caulines au niveau de la couronne externe et des fleurs dans la fleur. De manière générale, les pétales et étamines sont les organes les plus affectés, remplacés par des organes de type sepalloïdes à carpelloïdes (Fig.50A).

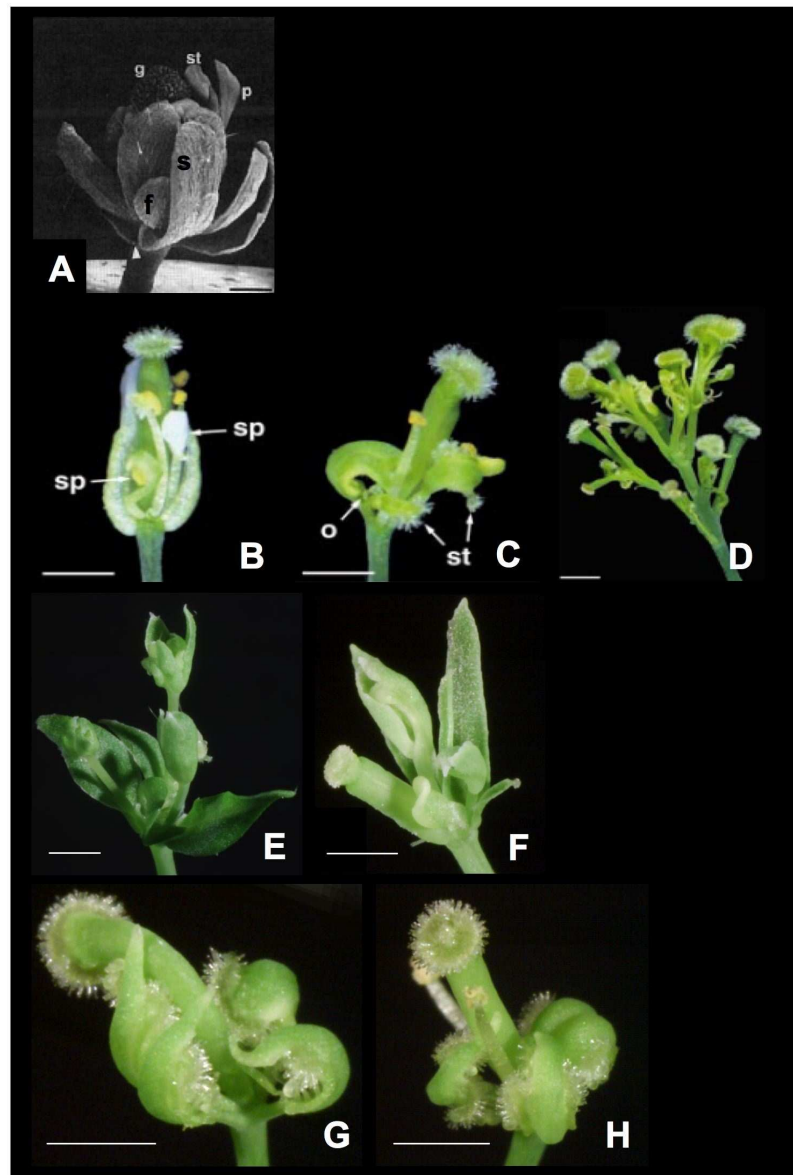


Figure 50: Mise en évidence d'une dimérisation de LFY chez *A.thaliana*

(A) Image de microscopie électronique à balayage d'une fleur *lfy-5* (Weigel et al., 1992). g=carpelle; st=étamine, p=pétale, s=sépale, f = primordium de fleur avorté; (B) Photographie d'une fleur à phénotype faible *LFY-VP16* (Parcy et al., 1998). o = ovule, sp = pétales, st = étamines; (C) Photographie d'une fleur à phénotype intermédiaire *LFY-VP16*; (D) Photographie d'une inflorescence à phénotype fort *LFY-VP16*; (E-F) Photographie de fleurs *lfy-5-VP16/lfy-5-VP16*, *lfy-12/lfy-12*; (G-H) Photographie de fleurs *lfy-5-VP16/lfy-5-VP16*, *LFY/LFY*. Les photographies sont faites à l'aide d'un microscope optique couplé à une caméra CCD. La barre d'échelle correspond à 0,1mm en A, 1mm en B-C-D-F-G-H et 5mm en E.

Les plantes *LFY-VP16* ont été décrites dans l'article de Parcy et collaborateurs (1998). Leurs phénotypes peuvent présenter différents degrés de sévérité, mais ils résultent tous d'une activation ectopique d'AG. Les phénotypes les plus faibles présentent des fleurs avec des pétales staminoïdes en deuxième couronne (Fig.50B). Les phénotypes intermédiaires présentent des sépales carpelloïdes, un nombre réduit d'étamines qui viennent prendre la place des pétales, et des carpelles surnuméraires (Fig.50C). Les phénotypes forts présentent quant à eux une structure carpelloïde principale, des couronnes surnuméraires, et une phyllotaxie en partie spiralée (Fig.50D). Il est important de noter ici que le phénotype *LFY-VP16* est étroitement dépendant du nombre de copies de *LFY* endogène (en

considérant 0 copie chez une plante *lfy/lfy*, 1 copie chez une plante *LFY/lfy*, et deux copies chez une plante *LFY/LFY*: moins il y a de copies de *LFY* endogène, plus le phénotype *LFY-VP16* est fort (Parcy et al., 1998).

Les fleurs *lfy-5-VP16/lfy-5-VP16*, *lfy-12/lfy-12* présentent un phénotype marqué par une couronne externe composée d'un ou plusieurs organes allongés qui s'assimilent à des organes de type feuilles caulines, un nombre faible voire souvent nul de pétales et étamines, et une grande fréquence de fleur dans la fleur (Fig.50E). Les fleurs se composent essentiellement d'organes de type sépale ou carpelle (Fig.50F). Toutes ces caractéristiques montrent donc que les plantes *lfy-5-VP16/lfy-5-VP16*, *lfy-12/lfy-12* présentent un phénotype mutant *lfy*. La fusion traductionnelle *lfy-5-VP16* est donc partiellement à totalement inactive, perte d'activité d'autant plus remarquée que normalement, *LFY-VP16* entraîne un phénotype *LFY-VP16* fort dans le fond *lfy*.

Dans le fond *LFY* homozygote sauvage, les fleurs *lfy-5-VP16/lfy-5-VP16* présentent un phénotype très sévère marqué par des structures composées uniquement de sépales carpelloïdes et de carpelles (Fig.50G). Des étamines se forment très occasionnellement entre ces organes (Fig.50H). Le phénotype des plantes *lfy-5-VP16/lfy-5-VP16*, *LFY/LFY* peut donc être assimilé à un phénotype de type intermédiaire à fort *LFY-VP16*. Chez ces plantes, il semble donc que la présence de protéines *LFY* endogènes ait rendue active la fusion traductionnelle *lfy-5-VP16*.

Les analyses des plantes *lfy-5-VP16* dans le fond *lfy* ou *LFY* sont donc cohérentes avec une possible interaction entre *lfy-5* et *LFY* endogène lors de l'activation des gènes cibles de *LFY*, résultats en accord avec les homodimérisations de *LFY* observées *in vitro* et chez la levure.

b) Mise en évidence d'une oligomérisation partiellement nucléaire par *Bimolecular Fluorescence Complementation* (BiFC)

1-Principe du BiFC

Le BiFC est une technique de visualisation d'interactions protéiques *in vivo* (Hu et al., 2002). Les deux protéines d'intérêt (la seule protéine *LFY* dans notre cas) sont fusionnées traductionnellement soit à la moitié N-terminale (YFP^N), soit à la moitié C-terminale (YFP^C) de la YFP (protéine fluorescente jaune). Si les deux protéines interagissent, le rapprochement physique des deux moitiés de la YFP va permettre la reconstitution d'une YFP fluorescente. Si les deux protéines d'intérêt n'interagissent pas, aucun signal YFP ne sera émis (Fig.51A).

Quatre constructions sont réalisées à partir de 4 vecteurs d'origine (BiFC1 à BiFC4). Dans le vecteur BiFC1, le gène d'intérêt, *LFY FL* ici, est inséré en position N-terminale de YFP^N (*LFY-BiFC1*). Dans le vecteur BiFC2, *LFY* est inséré en position C-terminale de YFP^N (*LFY-BiFC2*). Dans le vecteur BiFC3, *LFY* inséré en position C-terminale de YFP^C (*LFY-BiFC3*). Enfin, dans le vecteur BiFC4, *LFY* est inséré en position N-terminale de YFP^C (*LFY-BiFC4*; Fig.51B).

Dans le cas présent, l'expérience est réalisée par expression transitoire dans des feuilles de tabac et visualisée par microscopie confocale.

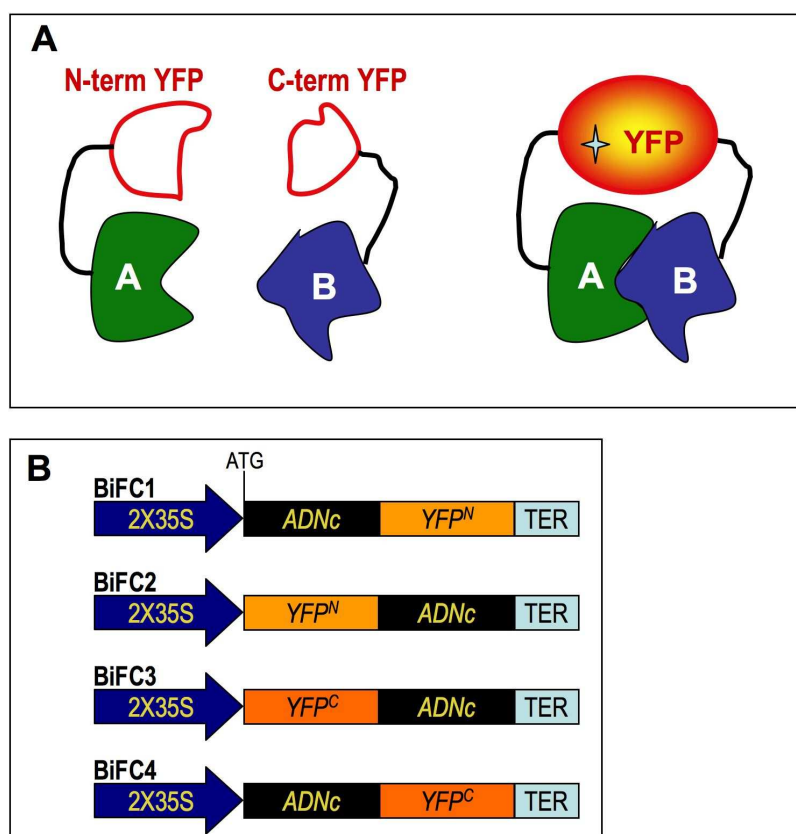


Figure 51: Principe du BiFC

(A) Représentation schématique des protéines A et B fusionnées à une demi YFP (portion N- ou C-terminale). L'interaction entre les protéines A et B conduit à la reconstitution d'une protéine YFP intègre; (B) Les quatre constructions BiFC. L'ADNc du gène d'intérêt est fusionné à l'extrémité 5' ou 3' de l'ADNc codant pour la moitié N-terminale de la YFP (respectivement BiFC1 et BiFC2), ou à l'extrémité 5' ou 3' de l'ADNc codant pour la moitié C-terminale (respectivement BiFC4 et BiFC3). De façon à tester l'homodimérisation de LFY, il faut pouvoir reconstituer la YFP. Nous réalisons donc les co-infiltrations LFY-BiFC1 avec LFY-BiFC3 ou LFY-BiFC4, et LFY-BiFC2 avec LFY-BiFC3 ou LFY-BiFC4.

2-Test d'oligomérisation de LFY par BiFC

Sur les quatre combinaisons possibles, seules LFY-BiFC2/LFY-BiFC4 et LFY-BiFC1/LFY-BiFC4 ont donné des signaux différents du bruit de fond (Fig.52B et D), nettement plus fort dans le cas de LFY-BiFC2/LFY-BiFC4 que de LFY-BiFC1/LFY-BiFC4. Ces signaux possèdent un pic d'émission à une longueur d'onde proche de la YFP, confirmant qu'il s'agit bien d'un signal BiFC. Ces observations suggèrent donc fortement que LFY est capable de s'oligomériser *in vivo*.

Pour confirmer ces résultats, plusieurs contrôles ont été réalisés. D'une part, toutes les constructions étudiées ont été infiltrées seules et n'ont donné aucun signal. D'autre part, nous avons utilisé une paire de facteurs de transcription à boîte MADS d'*A. majus*, DEFICIENS (DEF) et GLOBOSA (GLO), capables d'interagir entre eux et subséquentement d'entrer dans le noyau (mécanisme caractérisé à partir de l'étude de leurs orthologues AP3 et PI chez *A. thaliana*; McGonigle et al., 1996).

Alors que DEF et GLO donnent ensemble un signal de BiFC, aucun signal n'est visible lorsque la paire DEF-BiFC3/GLO-BiFC3 (YFP^C) est exprimée en même temps que LFY-BiFC2 (YFP^N) ou la paire DEF-BiFC2/GLO-BiFC2 (YFP^N) est exprimée en même temps que LFY-BiFC4 (YFP^C; images confocales similaires à celle reproduites en Fig.52A et C), témoignant de la spécificité des signaux LFY-BiFC2/LFY-BiFC4 et LFY-BiFC1/LFY-BiFC4.

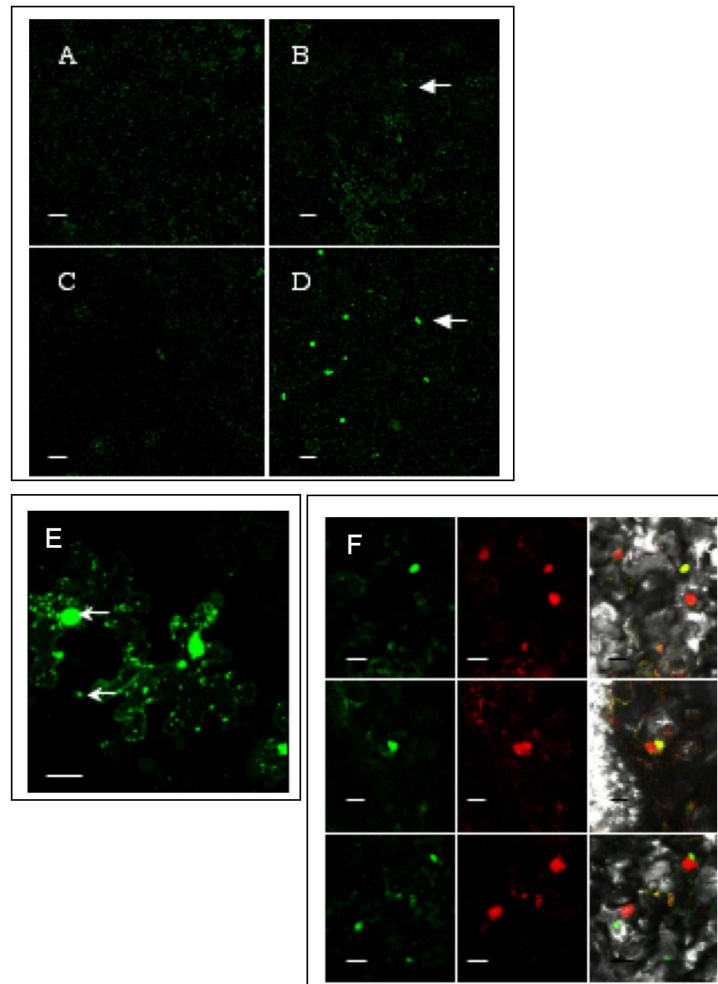


Figure 52: Mise en évidence d'une oligomérisation partiellement nucléaire de LFY par BiFC
 (A-D) Imagerie confocale des différentes combinaisons de LFY-BiFC: (A) LFY-BiFC1/LFY-BiFC3; (B) LFY-BiFC1/LFY-BiFC4; (C) LFY-BiFC2/LFY-BiFC3; (D) LFY-BiFC2/LFY-BiFC4. Les flèches indiquent la fluorescence YFP des signaux BiFC. La barre d'échelle correspond à 20 μ m; (E-F) Etude de la localisation subcellulaire du BiFC LFY: (E) Imagerie confocale de LFY-GFP. Les flèches pointent sur les deux types de compartiments d'adressage; (F) Imagerie confocale du BiFC LFY (colonne de gauche), du marqueur nucléaire (colonne centrale); et superposition (colonne de droite). La colocalisation donne une couleur jaune. Les trois lignes correspondent aux trois situations d'adressage de l'oligomère LFY observées. Gain 930 V pour le BiFC, et mRFP1, 650 V pour la GFP. La barre d'échelle correspond à 20 μ m.

3-Etude de la localisation subcellulaire de l'oligomère LFY

Pour déterminer la localisation subcellulaire de l'oligomère LFY, nous avons commencé par analyser la localisation de LFY à l'aide d'une fusion C-terminale (XW19(N)) à la protéine fluorescente verte complète (GFP; pic d'émission à 509 nm). LFY aurait deux sites d'adressage distincts. Le premier et principal est le noyau, en accord avec le rôle de facteur de transcription joué par LFY. Le second, qui

apparaît sous forme de ponctuations de tailles variables, correspondrait d'après Wu et collaborateurs (2003) aux plasmodesmes, ponts cytoplasmiques entre les cellules végétales (Fig.52E; Wu et al., 2003).

Pour déterminer où se trouve l'oligomère LFY, nous avons utilisé un marqueur nucléaire (l'histone H2B fusionnée à la mRFP1; Campbell et al., 2002), pour visualiser les noyaux grâce à leur fluorescence rouge. Trois situations ont alors été observées: i) l'oligomère LFY colocalise intégralement avec le marqueur nucléaire; ii) il n'occupe qu'une sous partie du noyau; iii) il est situé en périphérie du noyau (Fig.52F). Ces résultats sont donc en accord avec une oligomérisation de LFY en présence d'ADN (dans le noyau). Par contre, nous ne savons pas pourquoi dans certains cas tout le noyau est marqué, et dans d'autres qu'une sous partie du noyau. De plus, nous ne savons pas non plus si la présence de signaux BiFC en dehors du noyau a une signification biologique. Il pourrait s'agir d'agrégats conduisant à une interaction forcée des protéines LFY. Mais s'ils sont réels, ils signifient que LFY peut s'oligomériser en absence d'ADN. Dans quel compartiment cellulaire se situent alors ces signaux ? S'agit-il du réticulum endoplasmique par exemple, compartiment accolé au noyau ?

c) Mise en évidence d'une implication des résidus His387 et Arg390 dans la capacité de LFY FL à activer *AP1* par expression transitoire chez le tabac (travail d'E.Thévenon)

De la même manière qu'il l'a pratiqué chez la levure, Emmanuel Thévenon a étudié l'importance des résidus clés de la coopérativité His387 et Arg390 *in planta*. Il a pour cela co-exprimé transitoirement deux constructions au niveau de feuilles de tabac: i) une construction effectrice codant pour des protéines LFY, sauvage ou mutantes, fusionnées au domaine activateur constitutif VP16, dont l'expression est sous le contrôle du promoteur constitutif fort 35S; ii) une construction rapportrice où le gène *GUS* codant l'activité β -glucuronidase est placé sous le contrôle du promoteur d'AG. Les résultats, montrant une baisse significative de la capacité de LFY H387A/R390A à activer AG par rapport à LFY sauvage, viennent confirmer les tendances observées chez la levure, validant l'importance biologique des deux résidus His387 et Arg390 *in planta* (Fig.53). Plusieurs remarques sont tout de même à faire ici. Dans le cadre de cette expérience, de fortes variabilités sur les mesures ont été observées, affaiblissant la qualité des résultats. De plus, comme indicateur du 'bruit de fond', nous avons agroinfiltré la construction rapportrice seule correspondant à l'expression résiduelle d'AG. La mesure d'activité de la β -glucuronidase révèle une activité plus forte à partir de ce contrôle négatif qu'à partir de LFY H387A/R390A-VP16: non seulement cette dernière ne parvient pas à induire l'expression d'AG, mais elle fait même baisser le bruit de fond, ce qui semble difficile à expliquer.

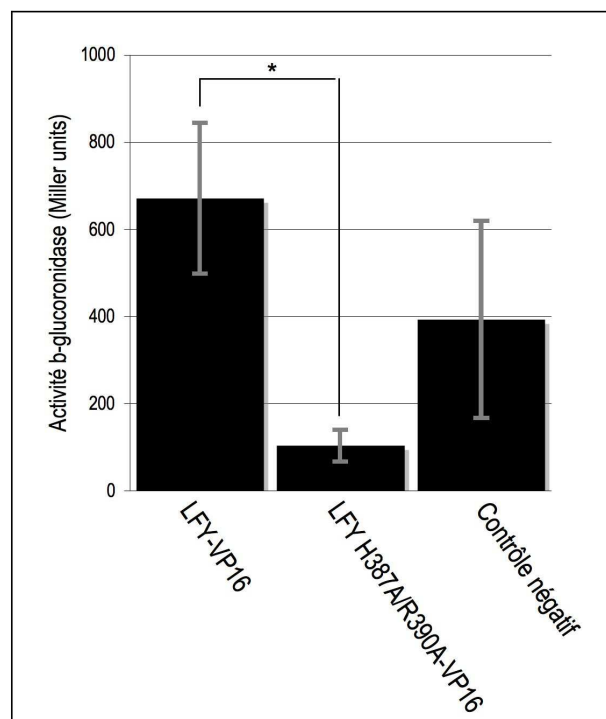


Figure 53: Mise en évidence d'une implication des résidus His387 et Arg390 dans la capacité de LFY FL à activer AG par expression transitoire chez le tabac *N.benthamiana* (travail d'E.Thévenon)

Etude quantitative de la capacité de LFY sauvage ou mutée H387A/R390A fusionnée au domaine activateur VP16 à induire l'expression de la construction rapportrice AG::GUS. Les protocoles relatifs à l'utilisation du gène rapporteur GUS sont précisés dans Parcy et al., 1994. Le contrôle négatif est obtenu à partir de l'expression du vecteur rapporteur seul (sans vecteur effecteur). Les barres d'erreur représentent les erreurs standard de la moyenne (SEM) N=10. Les astérisques signalent les écarts significatifs ($p < 0,05$). Deux expériences indépendantes ont été réalisées.

5. Discussion

a) Ce que l'approche biochimique nous a apporté

1-LFY adopte un comportement de liaison à l'ADN proche de celui de LFY-C

Les résultats préliminaires obtenus par approche biochimique montrent que LFY FL lie l'ADN sous forme d'un complexe majoritaire. Nous ne sommes pas sûrs de la nature de ce complexe, mais les données disponibles, à partir des analyses par EMSA avec AP1 WT, m1, m2 et m3, suggèrent qu'il pourrait s'agir d'un dimère comme il a été observé dans le cas de LFY-C. Les données obtenues à partir des EMSA réalisées avec les oligonucléotides AP1 mutés suggèrent également que ce dimère ne serait pas obligatoire puisqu'un complexe de plus faible masse parvient à se former, *a priori* composé d'un monomère LFY FL.

La protéine LFY FL serait donc capable de lier l'ADN sous forme monomérique et dimérique, le passage de cette première à cette seconde forme mettant probablement en jeu un mécanisme coopératif. Le K_d apparent de liaison de la protéine à l'oligonucléotide de référence AP1 a été estimé à 100nM. Enfin, la mutation de chaque demi-site d'AP1 (oligonucléotides AP1m1 et AP1m2) révèle une

spécificité de liaison pour chaque demi-site. Ces différentes données révèlent ainsi que LFY FL adopte un comportement de liaison à l'ADN similaire à celui de LFY-C, confirmant que le domaine conservé C-terminal joue un rôle central dans cette interaction, et que les données obtenues dans le contexte LFY-C semblent représentatives du comportement de la protéine complète. Il aurait en effet été possible que la région N-terminale de la protéine entière interfère au niveau de l'interaction du domaine C-terminal à l'ADN et modifie son affinité de liaison. Elle aurait également pu provoquer la dimérisation de LFY FL en l'absence d'ADN, et permettre à LFY de lier l'ADN en suivant la voie dimérique (Fig.43), remettant en cause la réalité de la coopérativité dans le contexte LFY FL. Dans ce cas cependant, on se serait attendu à ce que la double mutation H387A/R390A, compensée par la présence de la région N-terminale, ne modifie que très peu sa capacité à se lier à l'ADN, et qu'aucun monomère ne puisse lier seul l'ADN (comme dans le cas des sites *AP1* mutés dans un demi palindrome).

LFY FL se distingue en revanche de LFY-C par sa capacité à produire des complexes d'ordre supérieur sur l'ADN dès de faibles concentrations. Nous pouvons nous interroger sur la nature de ces complexes. Signalons qu'ils ont été observés avec une protéine LFY fusionnée à la GST, protéine capable de se dimériser et de perturber le comportement de liaison de la protéine d'intérêt (Niedziela-Majka et al., 1998). Peut-on ainsi incriminer la GST ? Cette différence pourrait aussi être attribuée au domaine N-terminal dont aucune fonction n'est connue à ce jour. La haute conservation des résidus 43 à 109 qui le composent implique très probablement un rôle important de cette région dans la fonctionnalité de LFY, rôle qui pourrait trouver raison tout ou partie dans la formation de ces 'gros' complexes. Il sera donc essentiel dans un futur proche de valider cette éventuelle implication du domaine conservé N-terminal et de déterminer quel peut être l'intérêt pour LFY de former de tels complexes. Ce dernier point sera discuté en discussion générale.

2-LFY pourrait contacter deux sites supplémentaires au niveau du promoteur d'*AP1*

L'analyse du promoteur d'*AP1* a permis de mettre en évidence deux séquences putatives de liaison de LFY à cette région d'ADN en plus de la séquence connue *AP1bs1*. Pour valider si LFY reconnaît effectivement ces séquences *AP1bs2* et *AP1bs3*, j'ai réalisé des EMSA et constaté que bien que LFY puisse contacter spécifiquement ces deux sites, l'affinité de ces interactions est très faible devant celle caractérisant la liaison de LFY à *AP1bs1*. Ces différences d'affinité pourraient traduire une inégalité fonctionnelle de ces sites *in planta*. Pour le savoir, Reyes Benlloch alors en stage post doctoral dans notre équipe a réalisé des lignées transgéniques où le gène rapporteur *GUS* a été placé sous le contrôle d'une sous partie du promoteur d'*AP1* contenant les sites *AP1 bs1*, *bs2* et *bs3* (*AP1::GUS*, promoteur de 600pb). Dans le contexte de ce promoteur, on détecte une forte activité *AP1::GUS* au niveau des jeunes méristèmes floraux (Fig.54A), témoignant d'une activation précoce normale d'*AP1*. En revanche, la délétion de *bs1* altère le niveau et le domaine d'expression d'*AP1* (Fig.54B). Les

mutations décrites *in vitro* d'*AP1 bs1* ('*AP1*') en *AP1m2* et *AP1m3* entraînent des défauts d'expression comparables à la délétion entière du site (Fig.54C et D). Ces résultats suggèrent donc que *bs1* joue un rôle fonctionnel important et que *bs2* et *bs3* ne sont pas suffisants pour compenser l'absence de *bs1*. Ils montrent aussi que les mutations *AP1m2* et *AP1m3* sont équivalentes biologiquement à la perte totale du site *bs1*.

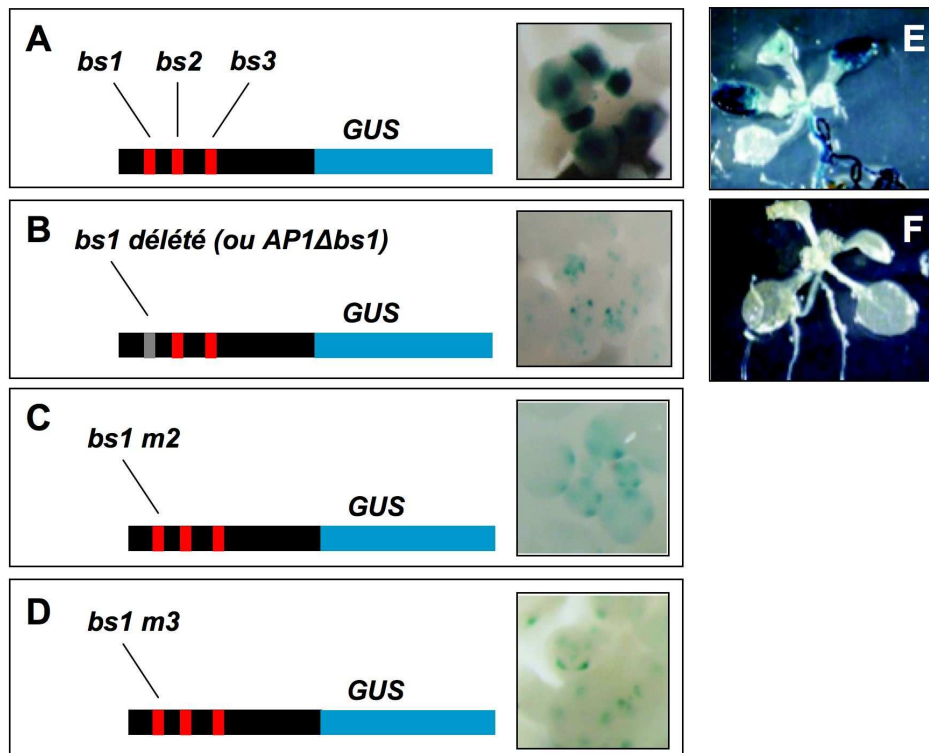


Figure 54: Etude de l'importance des sites *AP1bs1*, *bs2* et *bs3* *in vivo* (travail de R.Benlloch)
(A-D) Les constructions rapporteurs *AP1::GUS* sont schématisées à gauche, les photographies des colorations GUS correspondantes au niveau de jeunes stades de développements floraux sont illustrées à droite; (E-F) Photographies de jeunes plantules *HSP::LFY-VP16/AP1::GUS* (E) et *HSP::LFY-VP16/AP1Δbs1::GUS* (F) après traitement thermique.

Pour confirmer que l'importance du site *bs1* est bien dépendante de l'activité de LFY, Reyes a eu recours à des lignées stables d'*A. thaliana* co-exprimant une construction inductible à la chaleur de *LFY-VP16* (*HSP::LFY-VP16*) et une construction *AP1::GUS* délétée ou non au niveau du site *bs1*. Après son activation par traitement thermique, LFY est capable d'activer *AP1* (Fig.54E). Par contre, la délétion du site *bs1* annule toute expression d'*AP1*, suggérant fortement que l'activation de l'expression d'*AP1* via *bs1* est dépendante de LFY, et que *bs1* est le principal site de liaison de LFY au niveau du promoteur d'*AP1* (Fig.54F).

Ainsi, le fait que les mutations *AP1m2* et *AP1m3* aient une conséquence biologique si forte est particulièrement intéressant. Effectivement, nous avons vu *in vitro* que si *AP1m3* affecte considérablement l'affinité de liaison de LFY à l'ADN, *AP1m2* est encore reconnu par LFY-C selon une affinité environ 8 à 10 fois inférieure à celle de LFY-C pour *AP1 WT* (Fig.28C). Ce constat nous éclaire

donc sur la relation existante entre affinité de liaison à l'ADN et capacité de LFY à activer ses gènes cibles *in planta*.

b) Ce que les expériences *in vivo* nous ont apporté

1-LFY s'homodimérise *in vivo*

Les résultats obtenus dans la levure et la plante à partir de la coexpression de lfy5-VP16 et LFY de suggèrent que LFY est capable de s'homodimériser *in vivo*. D'après un essai de test en double hybride dans la levure, nous avons postulé que cette homodimérisation est dépendante de l'ADN. Récemment cependant, Chae et collaborateurs (2008) ont observé une homodimérisation LFY-LFY en double hybride classique, remettant en cause l'implication évoquée de l'ADN. La raison pour laquelle nous ne sommes pas parvenus dans nos conditions à reproduire ce résultat pourrait être liée à notre protocole. Les expériences de double hybride sont fortement dépendantes des natures du promoteur, des fusions traductionnelles (choix du domaine activateur et du domaine de liaison à l'ADN notamment), de la sensibilité du système rapporteur: beaucoup de paramètres peuvent influencer l'instauration d'une dimérisation entre les deux protéines cibles, de telle sorte qu'il est difficile de conclure sur la seule base d'un résultat de double hybride.

2-BiFC et mise en évidence d'une oligomérisation de LFY au niveau du noyau

Le BiFC offre la possibilité de tester *in planta* et selon un protocole rapide, l'interaction entre deux protéines. Toutefois, j'ai consacré un certain temps à optimiser le protocole car les signaux BiFC sont souvent très faibles comparativement à une protéine couplée à un fluorophore intègre, et sont de surcroît extrêmement dépendants de l'état physiologique de la plante. Aussi, j'ai délibérément choisi de travailler sur le tabac car j'ai constaté qu'en expression transitoire, le pourcentage de cellules transformées y est beaucoup plus important que chez *A. thaliana*, facilitant la co-expression de deux fusions au sein d'une même cellule (contrôle avec une construction 35S::GFP). D'un point de vue pratique, il est aussi plus facile de faire des agroinfiltrations sur des feuilles de tabac que d'*A. thaliana*.

Bien que travaillant *in vivo*, nous sommes conscients de ne pas être dans les conditions physiologiques où agit normalement LFY: les protéines de fusion LFY-BiFC ont été exprimées de façon transitoire dans des feuilles sous contrôle d'un promoteur constitutif fort. Nous pouvons donc nous interroger sur la réalité biologique de la localisation subcellulaire observée de l'oligomère. Nous avons vu qu'il avait deux régions d'adressage, à savoir le noyau et une région inconnue. Peut-on vraiment parler dans ce dernier cas de 'compartiment' ? J'ai évoqué la possibilité que cette région inconnue corresponde au reticulum endoplasmique, mais il se pourrait aussi qu'il s'agisse d'agrégats sachant que les constructions de BiFC ont été surexprimées. Les signaux BiFC observés au niveau du noyau constituent tout de même un indice intéressant en accord avec la possibilité que LFY se dimérise sur l'ADN.

3-Les résidus importants pour le mécanisme de coopérativité dans le contexte LFY-C *in vitro* interviendraient également dans l'activité de LFY *in vivo*

L'expérience de quantification de l'induction d'AG par les protéines LFY-VP16 sauvages ou mutées au niveau des résidus de la coopérativité suggère que l'His387 et l'Arg390 interviennent dans l'activité de LFY que ce soit chez la levure ou en expression transitoire chez le tabac.

Indiquons toutefois que ces résultats chez le tabac sont à prendre avec précaution étant donné la forte variabilité observée. Une confirmation sera nécessaire.

A l'avenir, il sera aussi nécessaire d'entreprendre cette série d'expériences dans le contexte du promoteur natif de LFY en transformation stable *in planta*, j'y reviendrai en discussion générale. Si la baisse de la capacité de LFY double mutant H387A/R390A à activer ses gènes cibles *AP1* et *AG* ne semble pas totalement abolie chez la levure, cet effet pourrait être beaucoup plus drastique chez *A. thaliana* dans le contexte du promoteur natif, comme le suggère le résultat préliminaire obtenu chez le tabac.

Bilan

Le bilan des publications à ce jour sur la protéine LFY établit que LFY lie directement et spécifiquement les régions régulatrices des gènes *AP1*, *AG* et *AP3*. Ces données permettent de proposer une séquence consensus de reconnaissance de l'ADN par LFY correspondant à CCANTG en considérant les sites reconnus au niveau du promoteur d'*AP1* et du second intron d'*AG*. Cette séquence se réduirait à CCNNNG en prenant également en considération les sites reconnus au niveau du promoteur d'*AP3* (Busch et al., 1999; Lamb et al., 2002; Parcy et al., 1998). La région conservée C-terminale de LFY assurerait une grande partie des fonctions de cette protéine selon des expériences de complémentation fonctionnelle *in planta* (Maizel et al., 2005). Soutenant cette proposition, cette région conservée C-terminale contiendrait le domaine de liaison à l'ADN (Maizel et al., 2005) et serait impliquée dans l'interaction avec le co-régulateur UFO (Chae et al., 2008). Hormis ces données, les connaissances relatives à la compréhension du mécanisme d'action de la protéine LFY étaient jusqu'ici très limitées.

Mon travail de thèse a permis de mieux comprendre le mode de fonctionnement de ce facteur de transcription unique. L'obtention de la structure tridimensionnelle du domaine LFY-C lié à l'ADN a confirmé l'originalité de cette protéine, spécifique au règne végétal, et a permis de disséquer les bases moléculaires de sa reconnaissance spécifique de l'ADN. En conjuguant ces résultats structuraux à une étude biochimique classique, ce travail a également permis de mettre en évidence une coopérativité de liaison de LFY sur l'ADN et de comprendre les bases moléculaires de ce mécanisme qui pourrait expliquer en partie la capacité de LFY à déclencher la transition florale. Au-delà de ces données, la structure de LFY-C liée à l'ADN représente un outil précieux qui facilitera l'identification des résidus ayant pu jouer un rôle dans l'évolution des fonctionnalités de cette protéine, de façon à comprendre comment LFY est intervenu dans l'apparition des plantes à fleurs (Frohlich, 2000).

1. Les données structurales sur le repliement de la protéine LFY

a) Un facteur de transcription original de la superfamille des protéines HTH

L'analyse par blast (serveur NCBI, serveur EMBL-EBI) de la séquence primaire de la protéine LFY n'avait révélé aucune homologie, hors de la famille LFY, avec les séquences des protéines enregistrées à ce jour sur les banques de données internationales, pas même au niveau de sous-domaines de la protéine. Par sa séquence, LFY semble tout à fait unique. Les données structurales confirment cette originalité au niveau de l'organisation tridimensionnelle de LFY. LFY est une protéine présentant un repliement à 7 hélices α d'un type nouveau. Toutefois, deux de ses hélices, les hélices $\alpha 2$ et $\alpha 3$, forment un motif de liaison à l'ADN extrêmement répandu de type HTH qui a déjà été identifié, par exemple, chez des acteurs du développement chez les animaux comme les protéines à

homéodomaine. Ce motif HTH se révèle donc favorable à la mise en place de mécanismes développementaux aussi bien chez les animaux que chez les végétaux.

b) Quelle est l'origine de *LFY* ?

D'où vient *LFY* ? Nous avons vu que *LFY* était présent chez toutes les plantes terrestres, depuis les Mousses. Est-ce que *LFY* est une innovation du règne végétal ? Il est difficile de répondre à cette question car aucune trace du gène n'est détectée chez les ancêtres des plantes terrestres, à savoir les algues. L'origine de *LFY* reste donc énigmatique: la copie du gène *LFY* pourrait avoir été perdue chez les descendants actuels des algues primitives, à moins qu'elle ait subi d'importantes modifications qui ne permettent plus de l'identifier. Une autre hypothèse serait que *LFY* soit le résultat d'un transfert horizontal à partir d'un gène du règne procaryote. Jusqu'ici, les similarités de séquence ne permettent pas d'affirmer qu'il existe un ancêtre commun avec les autres protéines HTH les plus proches, comme la Tc3 transposase, le domaine paired ou des protéines à homéodomaine (Fraenkel et al., 1998; van Pouderooyen et al., 1997; Xu et al., 1999), mais cette hypothèse reste envisageable. De futures données structurales sur de nouveaux facteurs de transcription apporteront peut-être des réponses à cette interrogation.

c) Une haute conservation de la protéine

LFY présente deux régions extrêmement bien conservées. La structure tridimensionnelle de l'une d'entre elle, la région C-terminale, révèle une organisation globulaire et des points de contact assurant la dimérisation de la protéine et sa liaison à l'ADN. Ces contraintes aussi bien internes qu'externes sur le repliement de la protéine, et par conséquent sur l'intégrité fonctionnelle de ce domaine, pourraient expliquer cette haute conservation de la région C-terminale. Des données fonctionnelles et structurales seraient nécessaires pour comprendre pourquoi la région N-terminale est elle aussi conservée.

d) L'ADN intervient-il sur la conformation de *LFY* ?

Les facteurs de transcription reconnaissent des séquences spécifiques au niveau de l'ADN. Pour sonder l'ADN jusqu'à atteindre leur cible, ils se déplaceraient par mouvement aléatoire brownien couplés à des épisodes de sauts anarchiques (Rezania et al., 2007; von Hippel and Berg, 1989). Une stratégie visant à accélérer cette prospection consiste à avoir une conformation permettant de limiter au maximum les liaisons de nature aspécifique et réciproquement avoir une conformation permettant d'augmenter l'affinité de leur liaison avec la séquence d'ADN reconnue spécifiquement. Pour ce faire, la plupart des facteurs de transcription changent de conformation pour se lier spécifiquement à leur ADN cible (von Hippel and Berg, 1989). Qu'en est-il dans le cas de *LFY* ? La structure du domaine *LFY*-C lié à l'ADN révèle deux régions extrêmement imbriquées dans l'ADN, à savoir l'Arg237, premier résidu ordonné N-terminal de la structure, contactant le petit sillon de l'ADN au niveau des bases ± 7 et

± 8 , et la Pro308 située à l'extrémité N-terminale de l'hélice $\alpha 3$, contactant le grand sillon au niveau des bases guanines ± 5 et ± 6 (Fig.35). Des structures de protéines à homeodomaine comme Engrailed montrent un changement de conformation du bras qui contacte le petit sillon en présence de l'ADN (Clarke et al., 1994). On pourrait imaginer que ce changement nécessite une séquence spécifique d'ADN et contribue donc à stabiliser le facteur sur sa cible. Il serait donc intéressant d'obtenir la structure de la protéine LFY-C en absence d'ADN. Les essais de cristallogénèse entrepris jusqu'ici sont restés vains. De plus, dans le cas de LFY, la dimérisation constitue un autre mécanisme qui pourrait s'ajouter au changement de conformation en stabilisant la liaison de chaque monomère à l'ADN, et qui augmente l'affinité de liaison de LFY à sa cible pseudo palindromique spécifique.

2. La coopérativité de liaison de LFY-C à l'ADN a-t-elle une réalité biologique?

La coopérativité de liaison de LFY-C à l'ADN a été mise en évidence par technique d'EMSA réalisée avec des oligonucléotides *AP1* et une gamme croissante de LFY-C faisant apparaître un complexe de deux protéines LFY par ADN toujours majoritaire par rapport à un plus petit complexe composé d'une seule protéine par ADN. De façon générale, cette coopérativité peut résulter de deux mécanismes distincts augmentant l'affinité de liaison du deuxième monomère à l'ADN, l'un mettant en jeu un changement conformationnel de l'ADN induit par la liaison du premier monomère, l'autre une dimérisation de la protéine sur l'ADN. Les données cristallographiques que nous avons obtenues ont montré que le mécanisme s'appliquant au cas de LFY-C était essentiellement une dimérisation de la protéine sur l'ADN impliquant deux résidus, l'His387 et l'Arg390. Ces données ouvrent une intéressante perspective d'étude car la coopérativité peut aider la mise en place de transitions développementales brutales et pourrait expliquer dans le cas présent la capacité de LFY à déclencher la transition florale. Il a été montré qu'une transition brutale demande de combiner coopérativité à une boucle d'autorégulation. La boucle d'autorégulation *via* l'activation d'*AP1* qui réactive LFY est connue (Liljgren et al., 1999). La combinaison de ces deux propriétés est sans doute importante pour expliquer comment LFY, dont l'expression croît de façon progressive, arrive à induire de façon brutale l'induction d'*AP1* et la transition florale. La question est désormais de valider si ce modèle a une réalité biologique, question qui s'applique à deux niveaux: i) la coopérativité s'observe-t-elle chez la protéine pleine longueur et ii) influe-t-elle sur la transition florale *in planta* ?

a) La coopérativité de liaison de LFY-C est-elle valable dans le contexte LFY FL ?

Des expériences préliminaires d'EMSA réalisées avec des oligonucléotides *AP1* et une concentration croissante en LFY-GST suggèrent que LFY FL serait capable de lier l'ADN de façon coopérative. En effet, de la même manière qu'il avait été observé avec la protéine LFY-C, la liaison de LFY-GST à *AP1 WT* met en évidence un petit complexe C1 de très faible intensité devant un complexe d'ordre supérieur C2. La proportion de complexe C1 augmente fortement suite à la mutation d'un demi-

site d'*AP1* (oligonucléotides *AP1 m1* ou *AP1 m2*). Ces résultats bien qu'encourageants exigent confirmation car ils ont été obtenus à partir d'une protéine couplée à la GST, étiquette capable de se dimériser et qui de ce fait pourrait biaiser le comportement de liaison de LFY FL. Il serait donc nécessaire de réitérer l'expérience avec une protéine LFY FL privée de l'étiquette GST ou avec une protéine mutée dans les résidus responsables de cette coopérativité potentielle His387 et Arg390.

b) La coopérativité intervient-elle dans l'activité de LFY FL *in vivo*?

Les résultats obtenus par BiFC *in planta* et par complémentation fonctionnelle avec le duo lfy5-VP16/LFY chez la levure et *in planta* montrent que LFY est capable de s'oligomériser *in vivo* et dans le noyau. En revanche, ces résultats ne donnent pas accès au nombre de protéines LFY par séquence cible d'ADN, et ne permettent pas de visualiser la protéine monomérique.

L'effet de la double mutation H387A/R390A sur la capacité de LFY FL à activer l'expression de ses gènes cibles *in vivo* a été étudié dans un second temps. Il est intéressant de noter que ces deux résidus His387 et Arg390 sont extrêmement bien conservés chez tous les Angiospermes dont l'ADNc de *LFY* a été cloné, pouvant constituer un indice en faveur d'un rôle lui-même conservé dans la dimérisation et par extension dans la coopérativité. En accord avec cette observation, nos résultats montrent aussi bien chez la levure qu'en expression transitoire dans des feuilles de tabac que cette double mutation occasionne une baisse d'efficacité de LFY-VP16 à activer AG, apportant un argument fort en faveur d'une importance de la dimérisation dans le contexte LFY FL *in vivo*.

Pour compléter cette caractérisation de la dimérisation de LFY à court terme, nous allons désormais étudier l'état d'oligomérisation de la protéine LFY FL libre par SEC ou MALLS-SEC. L'interface entre les deux monomères LFY-C est petite (420Å²), inférieure à de nombreux exemples décrits dans la littérature décrivant des valeurs d'interface de l'ordre de 1000Å² en moyenne (ex. de références bibliographiques avec des valeurs comprises entre 550 et 5400Å²; Baker et al., 2004; Bourne et al., 1996; Brown et al., 2007; Gouaux, 2004; Schrag et al., 2008; Yokoyama et al., 2008; Zimmer et al., 2006). Nous pourrions donc envisager que d'autres domaines de la protéine entière, comme le domaine conservé N-terminal, puissent augmenter cette interface et ainsi stabiliser la formation du dimère même en absence d'ADN. Nous allons aussi étudier l'effet de la double mutation H387A/R390A sur la capacité de LFY à s'oligomériser au niveau du noyau par BiFC chez le tabac.

A moyen terme, l'effet de la double mutation H387A/R390A sur la transition florale va être analysé à partir de plantes *A. thaliana* surexprimant une construction *LFY H387A/R390A* sous contrôle du promoteur constitutif fort 35S. Rappelons que les plantes 35S::*LFY* affichent un phénotype dominant de floraison extrêmement précoce (Weigel and Nilsson, 1995).

A plus long terme, nous allons réaliser des transformations stables de plantes où le gène *LFY* possédant la double mutation H387A/R390A est placé sous contrôle du promoteur natif *LFY*. Ces

dernières permettront une analyse plus fine et rigoureuse de l'effet de cette double mutation sur la transition florale.

L'étude de la dimérisation (et par extension de la coopérativité) *in vivo* s'avère une tâche délicate. En effet, la coopérativité conditionne fortement l'affinité de liaison de LFY à l'ADN. En mutant les deux acides aminés essentiels à cette coopérativité, nous allons affecter à la fois le mode de liaison de LFY mais aussi faire baisser son affinité. Ces deux caractéristiques sont difficiles à découpler.

Il serait possible d'annuler la supposée coopérativité en étudiant l'effet d'une dimérisation obligatoire de LFY (par création de liaison covalente entre les deux monomères), ce qui imposerait à LFY de lier l'ADN selon la voie dimérique (Fig.43). Cependant, comme je l'ai déjà mentionné précédemment, si dans l'absolu l'affinité de liaison est ainsi la même que celle d'une protéine utilisant la voie monomérique, cette situation ne se vérifie pas dans le contexte cellulaire. En effet, la voie dimérique en ralentissant la vitesse de prospection de l'ADN et en augmentant la probabilité de séquestrer le dimère au sein de complexes non spécifiques *in vivo* peut perturber indirectement l'affinité de liaison de LFY pour l'ADN. L'analyse est une nouvelle fois difficile: les observations obtenues seront-elles le résultat d'une perte de coopérativité ou d'une compétition trop forte par le reste du génome?

Un autre point important est à aborder ici. La coopérativité exige que LFY se dimérise sur l'ADN. Mais LFY agit-il systématiquement en dimère pour réguler l'expression de ses gènes cibles ? Nous verrons dans la partie traitant des corégulateurs qu'il serait possible que LFY et WUS s'hétérodimérisent au niveau des séquences régulatrices d'AG, hétérodimérisation inhibant la formation de l'homodimère LFY-LFY.

c) Quelle est la signification des complexes LFY/ADN d'ordre supérieur ?

Les expériences d'EMSA réalisées avec des oligonuléotides *AP1* et une gamme de concentrations en LFY FL suggèrent que LFY FL serait capable de lier l'ADN sous forme de monomère et dimère, mais serait également capable de former trois complexes d'ordre supérieur. Des complexes de taille plus élevée que le dimère ont également été observés avec LFY-C aux plus fortes concentrations utilisées. La question évidente qui s'impose ici est d'identifier la nature de ces complexes. Pour cela, il sera nécessaire en premier lieu de s'assurer que ces complexes ne résultent pas d'une oligomérisation non spécifique de LFY impliquant l'étiquette GST ou la GFP. Nous essaierons ensuite de déterminer la masse et par conséquent la nature de ces complexes par expériences GEMSA ou MALLS-SEC.

Ces complexes d'ordre supérieur alimentent des interrogations à deux niveaux: quelle(s) région(s) de LFY est/sont impliquée(s) dans la formation de ces complexes ? Quel est le rôle de ces complexes ?

Nous avons vu dans le cas de LFY-C qu'un de ces complexes d'ordre supérieur pouvait apparaître mais uniquement à très forte concentration protéique (de l'ordre de 5000nM contre 50nM

seulement dans le cas de LFY FL, fig.39 et 44). Pourquoi dans le cas de LFY FL voit-on apparaître ces différents complexes à si faible concentration? Est ce qu'une région de la protéine entière telle la région conservée N-terminale aux fonctions non établies pourrait intervenir dans la mise en place de ces complexes ? Des cas de tétramérisations mettant en jeu des dimérisations *via* la partie C-terminale et *via* la partie N-terminale ont d'ores et déjà été décrits dans la littérature (Leonard et al., 2004; Schreiter et al., 2003). Pour savoir si tel est le cas pour LFY, il serait intéressant de parvenir à produire le domaine N-terminal seul et à étudier s'il est capable de s'oligomériser.

Quelle pourrait être la signification biologique de ces complexes d'ordre supérieur ? Pourraient-ils permettre à LFY d'interagir avec des partenaires protéiques ? Ou pourraient-ils permettre à LFY d'occasionner un changement d'organisation de la structure secondaire de l'ADN ? Rappelons que LFY possède quatre sites de liaison au niveau des régions cis-régulatrices d'AG (*AG-I*, *AG-II*, *AG-III* et *AG-IV*). Nous pourrions imaginer que LFY par son oligomérisation ou par des liaisons avec des partenaires protéiques entraîne un repliement de l'ADN qui rapproche ces différents sites (Fig.55; Lamber et al., 2008; Yasmin et al., 2004). Cette éventualité pourrait s'appliquer également au cas d'*AP1*. Effectivement, nous avons mis en évidence que le promoteur d'*AP1* contenait lui aussi trois séquences putatives de liaison de LFY (*AP1 bs1*, *bs2* et *bs3*), les deux dernières étant des sites de plus faible affinité. Nous pourrions en effet concevoir que l'affinité de LFY pour les sites *bs2* et *bs3* augmente selon un comportement coopératif grâce à la présence à une distance déterminée d'un site de meilleure affinité comme *bs1*, possibilité déjà évoquée pour d'autres facteurs de transcription tels que la protéine TorR (Ansaldi et al., 2000). De plus, nous pouvons nous demander si la présence de plusieurs sites de liaison de LFY au niveau des séquences cis-régulatrices d'un même gène s'applique à d'autres gènes cibles de LFY, *i.e.* s'il s'agit d'un mécanisme d'action général de LFY ?

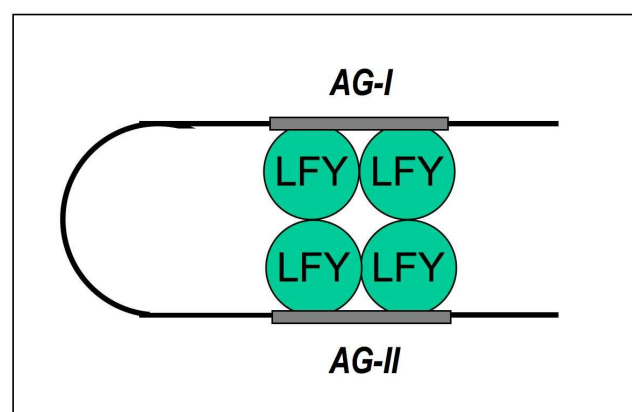


Figure 55: Modèle de liaison de LFY à deux séquences éloignées au niveau d'une même région cis-régulatrice

Figure ici l'exemple du second intron d'AG, avec deux sites de bonnes affinités, *AG-I* et *AG-II*. Nous pourrions envisager que cette situation s'applique également à des cas mettant en jeu des sites de forte et faible affinité comme par exemple au niveau du promoteur d'*AP1*.

Il serait intéressant de savoir si les gènes cibles de LFY possèdent systématiquement plusieurs sites de liaison. Afin d'identifier facilement ces sites sur la base de leur séquence, Edwige Moyroud, étudiante en thèse dans notre équipe, a élaboré une technique de sélection de sites de liaison ('selex') permettant de générer une matrice de liaison à l'ADN reconnue par LFY (Xue, 2005). A partir de ses résultats, il suffira de cribler les régions cis-régulatrices des gènes cibles de LFY afin d'identifier des sites potentiels de liaison et d'évaluer leur validité par approche biochimique classique.

Dans un second temps, nous comptons tester par technique de TPM (*Tethered Particle Motion*) si LFY peut effectivement modifier fortement la conformation de l'ADN au niveau de ces régions régulatrices (Nelson et al., 2006).

3. Caractérisation moléculaire de l'interaction LFY-C/ADN

L'interaction entre LFY et l'ADN au niveau des régions cis-régulatrices d'*AP1* et *AG* a tout d'abord été évaluée par EMSA et AF. Ces techniques ont donné accès aux K_d apparents de ces interactions, et l'AF nous a également permis d'estimer la constante de dissociation de l'interaction LFY/*AP1*. En jouant sur la taille et la séquence de l'oligonucléotide cible, ces techniques ont mis en évidence une liaison de LFY sur au moins 22pb, et une influence de cette séquence sur le comportement de liaison de LFY-C à l'ADN: LFY reconnaît des séquences palindromiques qu'il contacte en dimère selon un mode coopératif, coopérativité qui peut être sévèrement altérée par la mutation indépendante de chacun des demi-sites. L'obtention de la structure de LFY-C liée à *AP1* et *AG* a permis de disséquer les bases moléculaires de ces interactions en montrant que 12 acides aminés sont impliqués dans des contacts avec le squelette phosphate et trois autres acides aminés, l'Arg237, l'Asp291 et la Lys307, sont impliqués dans des reconnaissances base-spécifiques. Ces données structurales ont amené à compléter la séquence consensus CCANTG de liaison de LFY à l'*AP1* et *AG* sous la forme ${}^T/A\text{NNNNCCANTG}^T/{}_G\text{NNNN}^T/A$.

a) Détermination de l'affinité spécifique de liaison de LFY à l'ADN

Le K_d apparent de LFY pour trois de ses sites cibles, à savoir *AP1 bs1*, *AG-I* et *AG-II*, a été estimée à une valeur de l'ordre de 100 à 200nM. Cette affinité est dans la gamme des affinités décrites dans des interactions entre facteurs de transcription et leur cible d'ADN (Hurst, 1995) mais reste néanmoins relativement élevée.

b) Quelles séquences d'ADN sont reconnues par LFY ?

1-Réflexions autour des apports de la structure LFY-C/ADN

Bien que nous ayons essayé d'obtenir des cristaux de LFY-C avec des oligonucléotides pouvant atteindre 26pb, seuls ceux de 19pb ont donné une diffraction suffisante à la résolution de la structure du complexe LFY-C/ADN. Une expérience de compétition en AF entre un oligonucléotide de 30pb et des

oligonucléotides de 25 et 21pb avait pourtant indiqué que LFY contacte l'ADN au-delà de 21pb. Il est donc très possible que les données structurales soient partiellement incomplètes. Soutenant cette éventualité, la structure révèle des extensions basiques au niveau de LFY-C qui pourraient accueillir un fragment d'ADN plus long que celui ayant été résolu, mettant en jeu les lysines situées au niveau du court feuillet β N-terminal Lys253 et Lys254 (Fig.42). Du côté de LFY-C, la structure des premiers résidus à l'extrémité N-terminale du domaine conservé C-terminal n'a pu être résolue et révèle pourtant la présence d'une Arg235 (soit précédant de deux résidus l'Arg237) qui pourrait se situer à proximité immédiate de l'ADN. Pour étudier si effectivement les résidus Lys253, Lys254 et Arg235 participent à la liaison de LFY à l'ADN, nous pourrions tester l'effet de la mutation de ces résidus sur l'affinité de liaison de LFY à l'ADN. Cependant, cette technique ne nous donnerait pas accès à la nature des contacts, à savoir si ces résidus reconnaissent le squelette phosphate ou une base spécifique de l'ADN. En l'absence de données structurales, nous pourrions obtenir des indices à partir des données obtenues par selex qui pourrait laisser apparaître des bases conservées au niveau de la séquence consensus s'étalant au-delà de 21pb.

2-Questions auxquelles pourrait répondre le selex

Le selex pourrait apporter d'intéressantes données, nous nous limiterons ici à l'intérêt qu'il présente dans l'étude de LFY d'*A. thaliana* pour revenir à l'intérêt qu'il pourrait présenter dans une approche évo-dévo dans les points 5 et 6.

i) La principale portée du selex pourrait être de parvenir à identifier de **nouveaux gènes cibles de LFY**. En effet, cette technique permettra d'établir une matrice d'ADN (séquence de bases pondérées) reconnue par LFY, données qui pourraient permettre de cribler le génome d'*A. thaliana* pour y rechercher des sites de liaison et confronter la liste des gènes qui en possèdent aux cibles potentielles de LFY (Schmid et al., 2003; Wagner et al., 2004; William et al., 2004). Cette étude pourrait aussi permettre de savoir comment sont disposés les sites de liaison des gènes cibles (nombre, espacement) et si la coexistence de sites de forte et faible affinité est souvent observée. L'ensemble de ces travaux permettrait de disposer de règles pour la recherche de gènes cibles sur la seule base de leur séquence.

ii) La confrontation des listes de gènes cibles de LFY (Schmid et al., 2003; Wagner et al., 2004; William et al., 2004) avec la présence de site de liaison dans leurs séquences régulatrices devrait permettre de savoir s'il existe des gènes cibles directs sans site de liaison pour le dimère (deux demi-sites juxtaposés), suggérant que LFY pourrait aussi réguler des gènes sous forme de monomère ou d'hétérodimère. En effet, nous avons vu que LFY se liait en dimère sur chaque demi-palindrome d'une séquence palindromique, mais qu'il était également capable de lier chaque demi-site intègre sous forme monomérique (expériences réalisées avec des oligonucléotides mutés dans un demi-palindrome). Ainsi,

serait-il possible que LFY ne reconnaisse qu'un 'demi-site' au niveau de certains de ses gènes cibles, *i.e.* **qu'il lie les régions cis régulatrices de certains de ses gènes cibles uniquement sous forme monomérique** ? Si tel est le cas, le selex pourrait mettre en évidence des séquences ne présentant qu'un demi-site. Signalons toutefois que le selex tend à sélectionner les sites de très bonne affinité. Sachant que le monomère a théoriquement une affinité de liaison à l'ADN plus faible que celle du dimère, il sera sans doute nécessaire dans ce cas de limiter autant que possible le processus de sélection tout en obtenant des séquences reconnues spécifiquement par LFY.

iii) Dans les différents cas étudiés (*AP1*, *AG* et *AP3*; Busch et al., 1999; Hong et al., 2003; Lamb et al., 2002; Parcy et al., 1998), tous les sites reconnus par LFY présentent deux demi-sites toujours espacés par une **seule et unique base** (correspondant à la base médiane N de la séquence consensus CCANTGT). Les données structurales laissent supposer que cet espacement déterminé est sans doute le mieux adapté au bon positionnement des deux monomères l'un par rapport à l'autre, et que modifier cet espacement entraverait la formation du dimère et donc la coopérativité de liaison. Cette hypothèse va être testée par EMSA avec des oligonucléotides présentant aucun N (CCATG) ou 2N (CCANNTG).

4. Comment LFY agit-il avec ses corégulateurs ?

LFY, pour assurer son rôle de gène FMI, s'exprime très précocement au niveau de la mise en place des futures fleurs. L'analyse de son patron d'expression révèle que ses ARNm sont présents dès le stade *anlagen* (ou stade floral 0) sur l'ensemble du futur méristème floral, patron d'expression qui reste uniforme jusqu'au stade de développement floral 3 (Blazquez et al., 1997; Weigel et al., 1992). Ainsi, comment LFY peut-elle réguler l'expression des gènes homéotiques floraux à des territoires confinés du méristème floral ? Cette question a trouvé réponse dans l'identification de co-régulateurs de *LFY* dont le patron d'expression confère une délimitation spatiale à l'activité de *LFY*. En effet, UFO permet à LFY d'activer *AP3* au niveau des futurs pétales et étamines, et WUS permet à LFY d'activer *AG* au centre du méristème floral, en place des futurs étamines et carpelles. Nous allons voir que l'obtention de la structure de LFY-C à l'ADN, et nos données nouvelles de biochimie, nous a amené à revisiter ces corégulations.

a) Cas de la régulation d'*AG* par LFY et WUS

Lohmann et collaborateurs ont identifié le corégulateur WUS en 2001 et montré qu'il agissait de façon synergique avec LFY pour activer *AG*, chacune de ces protéines n'étant pas capable d'activer *AG* à elle seule. *WUS* code pour une protéine à homéodomaine capable de lier l'ADN sur une séquence consensus putative propre à cette famille de facteur de transcription de la forme TTAATGG (Gehring et al., 1994). Cette séquence est observée à proximité des deux sites de liaisons de LFY décrits jusqu'alors au niveau du second intron d'*AG* (*AG-I* et *AG-II*, Fig.14). Dans un premier temps, en

observant que les mutations, soit des sites de liaison de WUS, soit des sites de liaison de LFY, annulent la capacité du duo LFY/WUS à activer AG par expérience des transactivation chez la levure, ces auteurs déduisent que LFY et WUS contactent le deuxième intron d'AG en même temps. La mutation d'un seul des deux sites de liaison de LFY n'a en revanche qu'un effet modéré. Par EMSA, ces auteurs montrent ensuite que, malgré la proximité des sites de liaison de LFY et WUS, le mélange de ces deux protéines en présence d'ADN n'engendre pas la formation d'un complexe supplémentaire: LFY et WUS contacteraient indépendamment l'ADN. Depuis ces données, deux nouveaux sites de liaison putatifs de LFY ont été identifiés, *AG-III* étant lui aussi couplé à un site putatif de liaison de WUS (Hong et al., 2003).

Lorsque nous avons obtenu la structure de LFY-C lié à *AG-I*, nous avons regardé comment pouvait se positionner WUS, mais nous nous sommes rapidement heurtés à un problème: au moment de la parution de l'article de Lohmann et collaborateurs, la séquence consensus de LFY admise était de la forme CCANTG, ignorant que deux protéines LFY lient simultanément l'ADN sur une séquence bien plus longue. Dans le cas de WUS, il serait possible également que cette protéine s'homodimérise comme il l'a été observé par double hybride avec l'orthologue de WUS chez le riz, bien que ces travaux ne permettent pas de conclure si cette homodimérisation peut s'effectuer sur la séquence cible d'ADN de WUS (Nagasaki et al., 2005). En revisitant le positionnement des sites de liaison de LFY et WUS, nous avons donc constaté que les deux sites couvraient partiellement la même région d'ADN (Fig.56A). Il ne semble donc pas possible physiquement que l'homodimère LFY contacte l'ADN simultanément à WUS au niveau de deux séquences de liaison LFY/WUS adjacentes, si ce n'est si le dimère LFY contacte une région donnée de l'ADN et WUS la région diamétralement opposée (Fig.56B). Pour tester cette éventualité, François Parcy a réalisé la modélisation structurale de LFY et WUS au niveau de leur séquence cible (Fig.56C) et constaté que le modèle de liaison simultanée du dimère LFY et de WUS décrit en Fig.56B (à gauche) semble peu probable. Ce constat nous a amené à avancer une nouvelle hypothèse. Pour concilier l'action synergique avec les données de structure, nous proposons donc que LFY et WUS puissent former un hétérodimère sur AG (Fig.56B à droite), ce qui de plus expliquerait d'un point de vue moléculaire pourquoi des sites de liaison de WUS sont observés si fréquemment à proximité de sites de liaison de LFY. Cette hypothèse conduit à une autre question. Les sites *AG-I*, *AG-II* et *AG-III*, tous trois composés d'un site de liaison de LFY couplé à un site de liaison de WUS, pourraient ne pas avoir un rôle fonctionnel équivalent. En effet, si le site de liaison de WUS est situé en 5' des sites de liaison LFY au niveau *AG-I* et *AG-III*, il est situé en 3' du site au niveau d'*AG-II*: l'orientation relative des sites de liaison respectifs de LFY et WUS n'est pas équivalente entre *AG-I*, *AG-II* et *AG-III* (Fig.14). La question revient donc à savoir si des interactions LFY/WUS peuvent mettre en jeu différents points de contact de part et d'autre de chaque protéine comme l'illustre le modèle en figure 56D ? Ou ces trois sites sont-ils contactés différemment par LFY et WUS? La synergie

d'activation observée chez la levure par Lohmann et collaborateurs a été obtenue à partir d'un fragment d'ADN contenant les sites *AG-I* et *AG-II*, tandis que les EMSA décrivent la liaison de LFY et WUS au niveau du site *AG-II* uniquement. Il serait donc intéressant de coupler ses EMSA à une analyse avec *AG-I* également.

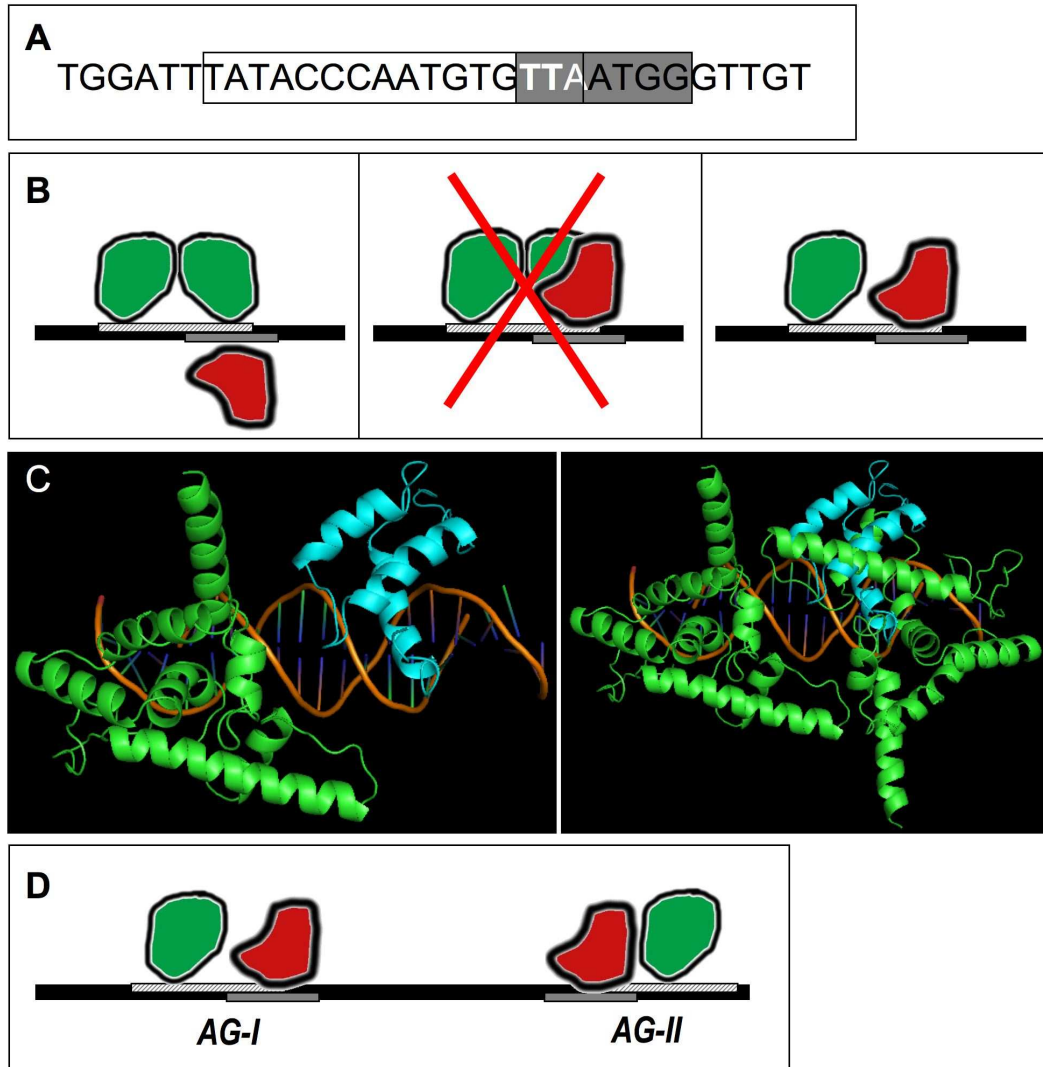


Figure 56: Liaison de LFY et WUS au niveau d'AG

(A) Représentation du site *AG-II*. En encadré blanc, séquence putative reconnue par le dimère LFY ne prenant en compte que les données apportées par la structure ($^A/TNNNNCCANTG^T/GNNNN^T/A$); en encadré gris, séquence consensus de liaison de WUS; la base thymine mentionnée en blanc est partie intégrante des séquences reconnues à la fois par LFY et WUS; (B) Modèle de liaison simultanée du dimère LFY et de WUS au niveau d'*AG-I*. Le site de liaison de LFY à l'ADN est schématisé en rayé gris, le site de liaison de WUS en gris, les protéines LFY sont en vert et la protéine WUS en rouge; le modèle propose que selon la position de WUS sur l'ADN, l'encombrement stérique permette (à gauche) ou ne permette pas (au milieu) une coexistence avec le dimère LFY; l'hétérodimère LFY/WUS (à droite) est également envisageable; (C) Modélisation à gauche de l'hétérodimère LFY/WUS au niveau d'*AG-I*, et à droite de l'homodimère LFY lié en même temps que WUS sur *AG-I* à l'aide du logiciel PyMOL. Le modèle structural de WUS a été établi à partir de la structure d'une protéine à homéodomaine homologue; (D) Modèle de liaison de l'hétérodimère LFY/WUS aux sites *AG-I* et *AG-II*.

Pour tester l'hypothèse d'une action en hétérodimère, Emmanuel Thévenon, technicien dans notre équipe, a testé par expérience de transactivation qualitative chez la levure la capacité de LFY et WUS à activer AG (Fig.57). Après avoir confirmé que la co-expression de LFY et WUS activait bien l'expression d'AG, Emmanuel a testé l'effet d'une délétion de LFY au niveau de la région non conservée C-terminale (LFY Δ , résidus 391-424). Cette région n'a pas de rôle démontré: elle n'apparaît pas dans la structure de LFY-C liée à l'ADN suggérant qu'elle n'est pas structurée, et ne semble pas intervenir dans la liaison à l'ADN. Elle ne semble pas non plus intervenir dans la coopérativité comme le suggère le comportement de liaison des protéines produites à partir de pCH43 (résidus 224-391), protéines délétées de cette même région dans le contexte LFY-C. En revanche, dans son expérience, Emmanuel constate que cette truncation annule la synergie d'activation LFY/WUS (Fig.57): la région C-terminale non conservée pourrait donc intervenir dans une putative hétérodimérisation LFY/WUS. Pour affiner ces résultats, Emmanuel est en train de confirmer cette expérience de transactivation chez la levure par des mesures quantitatives. Il a pour l'instant mis au point un protocole permettant de visualiser la synergie d'activation LFY/WUS et devrait prochainement tester la capacité de LFY Δ /WUS à induire AG.

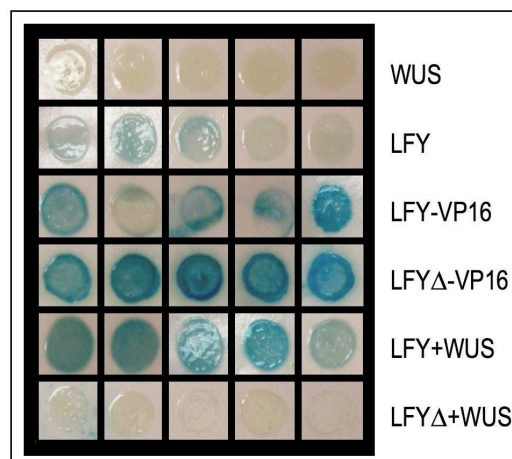


Figure 57: Etude de la synergie d'activation d'AG par LFY/WUS chez la levure (travail d'E. Thévenon)

Etude qualitative de la capacité de différentes protéines LFY et/ou WUS à induire l'expression du gène rapporteur *LacZ* placé sous le contrôle du promoteur d'AG par expérience de transactivation. L'activité est visualisée par suivie de la coloration bleue du produit de la réaction de catalyse du substrat X Gal par la β -galactosidase. LFY Δ correspond à la protéine LFY FL délétée des résidus 391 à 424.

Dans l'optique de tester par d'autres stratégies cette interaction LFY/WUS potentielle, nous allons désormais étudier selon cette même approche dans la levure la synergie d'activation de WUS et LFY mutée dans les résidus His387 et Arg390, impliqués dans l'homodimérisation de LFY. Nous avons vu que cette double mutation empêche LFY couplée au domaine activateur constitutif VP16 d'activer AG, alors que la protéine sauvage couplée à ce même domaine VP16 active AG. Ainsi, si la synergie n'est pas affectée, on peut conclure qu'elle ne fait pas intervenir l'homodimère LFY et que les 2 résidus ne sont pas impliqués dans un éventuel contact avec WUS. Par contre, si elle est affectée, on ne pourra rien conclure.

J'ai essayé pour ma part d'étudier la liaison de LFY et WUS à AG par approche biochimique. J'ai pour cela essayé de produire WUS à partir de différents vecteurs (décrits en Matériel et méthodes) en bactéries *E. coli* issues de différentes souches. Malgré les tentatives d'optimisation des conditions d'induction, l'expression de WUS reste faible et exige de nouveaux essais. Des expériences préliminaires à partir de protéines très impures (dont la présence a été confirmée par Western Blot avec des anticorps anti-WUS) n'a pas permis de mettre en évidence d'interaction spécifique à l'ADN par EMSA. Pour pallier à ce problème de quantité et qualité de la protéine, j'ai entrepris selon le même principe que celui appliqué à LFY l'étude d'un sous-domaine de WUS, à savoir l'homéodomaine (WUS-Hd) qui d'après les connaissances sur les protéines à homéodomaine serait le domaine de liaison à l'ADN de WUS (Luscombe et al., 2000). WUS-Hd a pu être très bien exprimé et purifié jusqu'à un haut niveau de pureté (Fig.58A et B). Malheureusement, un test par EMSA de liaison de cette protéine aux oligonucléotides d'AG-I, AG-II et à un oligonucléotide AP1 ne contenant pas de séquence prédite de liaison de WUS (contrôle négatif) n'a pu mettre en évidence de liaison spécifique à faible concentration protéique. Seules des bandes très étalées ('smear') sont observées avec 500nM de WUS-Hd, quelque soit l'oligonucléotide considéré. Il semble néanmoins qu'elles soient plus prononcées à partir des deux oligonucléotides AG qu'à partir d'AP1. J'ai également essayé de mélanger WUS-Hd à LFY-C en présence de ces différents oligonucléotides dans l'éventualité que la présence de LFY stabilise WUS mais aucune expérience n'a laissé apparaître de retard spécifique au mélange de ces deux protéines (Fig.58C). Il m'est difficile de tirer une conclusion de ces résultats. Les conditions de liaison appliquées ici sont les mêmes que celles pratiquées par Lohmann et collaborateurs, conditions qui ont permis de mettre en évidence des interactions entre WUS et l'oligonucléotide AG-II. Cet homéodomaine semble bien structuré d'après la littérature sur ce type de domaines et selon des logiciels de prédiction (Fig.58D), et il semble peu probable qu'il ne soit pas impliqué dans la liaison de WUS à l'ADN. Pourquoi alors le domaine WUS-Hd n'est-il pas capable de reconnaître ses séquences cibles ? Peut-on mettre en cause l'affinité de liaison de cet homéodomaine à l'ADN ? Peut-être en effet qu'une autre région de WUS contribuerait également à l'interaction WUS/ADN. Peut-on incriminer la quantité de compétiteurs aspécifiques d'ADN ? Dans cette présente expérience d'EMSA, nous avons mis une quantité d'ADN compétiteur froid établie à partir de la liaison de LFY à l'ADN, quantité suffisamment élevée pour abolir totalement la liaison aspécifique de LFY pour l'ADN. Dans le cas de WUS, les affinités aspécifique et spécifique pourraient être d'un autre ordre de grandeur. Un autre point potentiellement important est à soulever ici : dans leur article, Lohmann et collaborateurs effectuent leurs EMSA à partir de protéines produites chez la levure, tandis que je produis WUS-Hd en bactéries. La protéine WUS nécessiterait-elle pour contacter l'ADN de subir des modifications post-traductionnelles telles que des glycosylations observées uniquement chez les eucaryotes ?

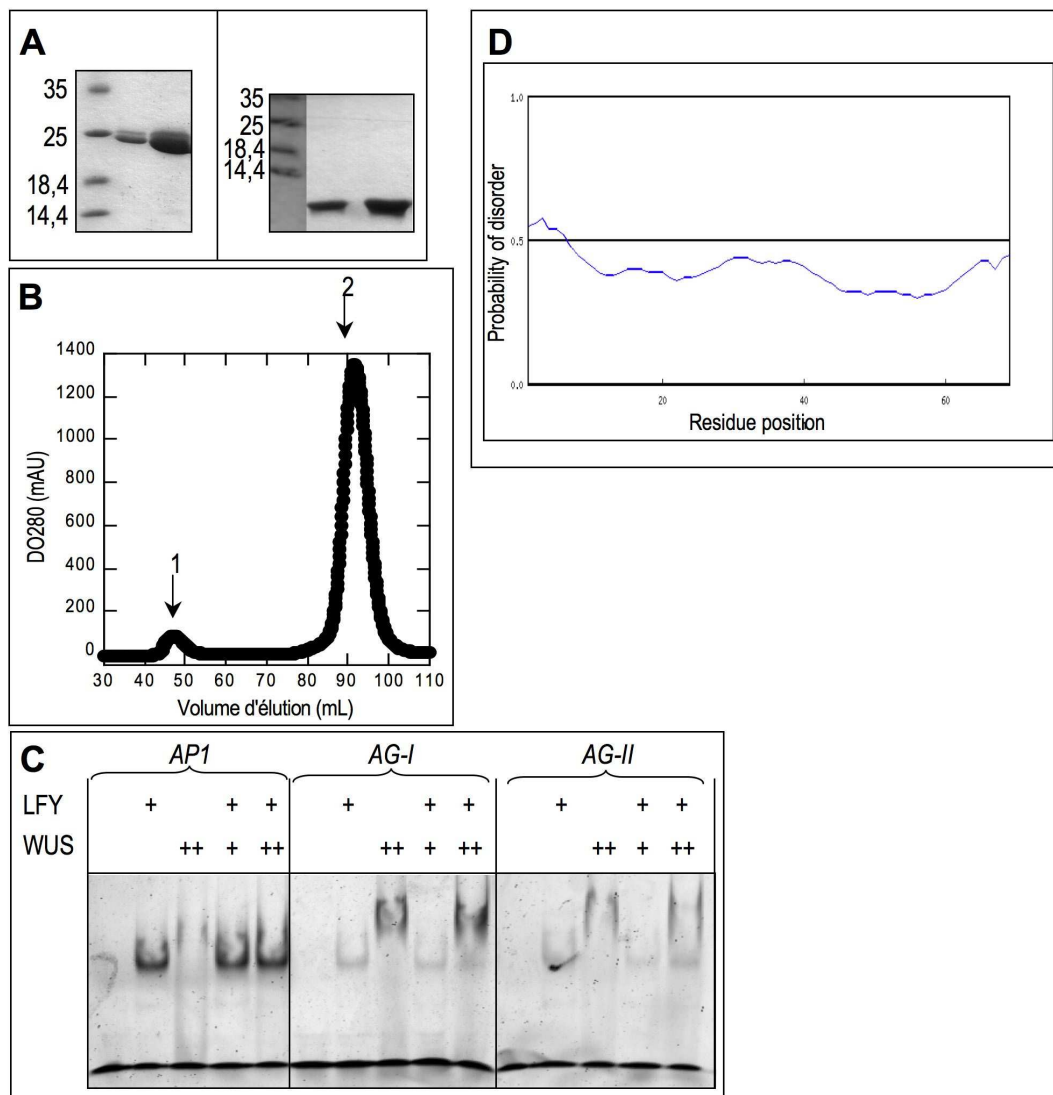


Figure 58: Etude du domaine WUS-Hd

(A) Analyse par gel SDS-PAGE de la protéine WUS-Hd avec étiquette thioredoxin-6his (à gauche, 1 et 5 μ L déposés) ou clivée de son étiquette de solubilisation/purification (à droite, 0,5 et 1 μ L déposés). La gamme obtenue avec les protéines standard est signalée à gauche (avec masse moléculaire en kDa) ; (B) Profil de SEC (DO₂₈₀) de WUS-Hd (la flèche 1 signale la fraction agrégée, la flèche 2 la fraction non agrégée) ; (C) EMSA réalisée avec 10nM d'oligonucléotides marqués (précisé), 500nM (+) LFY-C, et 500nM (+) ou 5000nM (++) WUS-Hd. Les bandes très étalées visibles à 5000nM WUS ne sont pas observées à de plus faibles concentrations; (D) Exemple de prédiction du désordre intrinsèque du domaine WUS-Hd à l'aide du logiciel RONN (*Regional Order Neural Network*; http://www.strubi.ox.ac.uk/cgi-bin/disorder_results_jan2005.cgi) de la division de biologie structurale de l'Université d'Oxford (STRUBI). Les régions structurées correspondent aux zones où la probabilité de désordre est inférieure 0.5.

Le mécanisme moléculaire à la base de la synergie d'activation d'AG exercée par LFY et WUS reste donc incompris. La seule donnée publiée à ce jour par approche biochimique sur cette régulation, l'expérience d'EMSA de Lohmann et collaborateurs, ne laisse pas apparaître de LFY et WUS de retard autre que ceux résultants de chaque protéine liée seule à AG-II. Sachant que la migration n'est pas uniquement dépendante de la masse des molécules mais également de leur conformation et de leur point isoélectrique, il serait possible que l'hétérodimère migre dans ces conditions d'expériences de la même manière que le complexe formé par LFY ou WUS seule. Mentionnons aussi que cette expérience

ne prend pas en compte la concentration ni l'affinité de liaison à l'ADN des deux protéines, laissant envisager que les quantités de protéines utilisées pour cette expérience ne soient pas favorables à l'apparition de l'hétérodimère.

Des alternatives à l'hypothèse d'hétérodimérisation peuvent être avancées. LFY et WUS pourraient contacter des sites différents au niveau du second intron d'AG: si l'homodimère LFY contacte AG-I, WUS contacterait AG-II et inversement (modèle auquel il pourrait être inclus AG-III également). Cette redondance fonctionnelle expliquerait pourquoi la mutation d'un seul site LFY n'a que peu d'incidence sur la capacité du duo LFY/WUS à activer AG chez la levure (Lohmann et al., 2001). Rien ne s'oppose non plus à l'hypothèse que l'action de LFY et WUS soit régulée dans le temps, LFY contactant ses séquences cibles en premier puis laissant place à WUS ou inversement.

A long terme, nous espérons trouver réponse à cette énigme par des études biochimiques combinées à une éventuelle cristallisation du complexe WUS/AG ou LFY/WUS/AG, sous condition que les deux protéines coexistent à un moment donné sur AG. Des mutations ponctuelles au niveau des résidus de LFY nécessaires à son interaction avec WUS seront alors introduites chez *A. thaliana* dans le fond mutant *lfy* afin de valider la relevance biologique de ces données. D'ici là, nous envisageons d'introduire *in planta* une version de LFY délétée de sa région hydrophobe C-terminale (résidus 391-421), expérience qui pourrait venir compléter les résultats obtenus chez la levure.

b) Cas de la régulation d'AP3 par LFY et UFO

L'importance de LFY et UFO dans l'activation de l'expression d'AP3 a été mise en évidence par approche génétique chez *A. thaliana* (Lee et al., 1997b; Parcy et al., 1998). Lamb et collaborateurs en 2002 ont montré par EMSA (Fig.13D) que cette activation passait par une liaison directe de LFY au promoteur d'AP3. Très récemment, Chae et collaborateurs (2008) ont décrit une série d'expériences faisant état d'une interaction physique de LFY et UFO indépendante de l'ADN, mais qui permettrait à LFY de recruter UFO au niveau du promoteur d'AP3. En effet, par EMSA, ces auteurs montrent qu'à elle seule UFO n'est pas capable de lier AP3, mais que UFO provoque un « *supershift* » du complexe LFY/ADN, suggérant qu'il s'y incorpore (Fig.13F). Cette interaction entre LFY et UFO a été confirmée par expérience de double hybride chez la levure, expérience réalisée avec des protéines LFY entières ou tronquées qui suggère que ce contact impliquerait la région conservée C-terminale de LFY.

La publication de Chae et collaborateurs apporte d'autres éléments de réflexion. UFO est une protéine à boîte-F, famille de protéines très souvent décrites en tant que composants de complexes SCF. Elles agiraient en tant qu'E3 ubiquitine ligase pour étiqueter et adresser des protéines vers la voie de dégradation dépendante du protéasome (Deshaies, 1999). Pour tester si une telle fonction d'UFO intervient dans sa capacité à réguler AP3, Chae et collaborateurs apportent des résultats préliminaires sur une possible ubiquitination de LFY qui impliquerait partiellement UFO, et montrent que l'activité du protéasome est nécessaire à la régulation d'AP3.

Ces résultats apportent une base intéressante à la compréhension du mécanisme moléculaire régissant la régulation d'AP3 exercée par LFY et UFO. Nous souhaiterions la compléter sur plusieurs points à commencer par la cartographie des points de contact entre LFY et UFO. Si Chae et collaborateurs (2008) ciblent le domaine C-terminal de LFY, Ronald Koes (Université de Vrije, Amsterdam, Pays-Bas) qui étudie cette même interaction chez le pétunia nous a confié qu'il possédait des résultats impliquant au contraire le domaine N-terminal de LFY (thèse d'Alexandra B. Rebocho, 2008). Afin de trancher sur cette ambiguïté, Edwige Moyroud a tout d'abord confirmé la liaison de LFY à AP3 en travaillant à partir d'un oligonucléotide centré sur la séquence consensus de LFY. Alors que l'oligonucléotide dessiné par Lamb et collaborateurs (2002) est reconnue avec une mauvaise affinité par LFY, cet oligonucléotide dessiné par Edwige montre que LFY reconnaît ce site selon une affinité d'un même ordre de grandeur que celle de LFY pour AP1 (Fig.59).

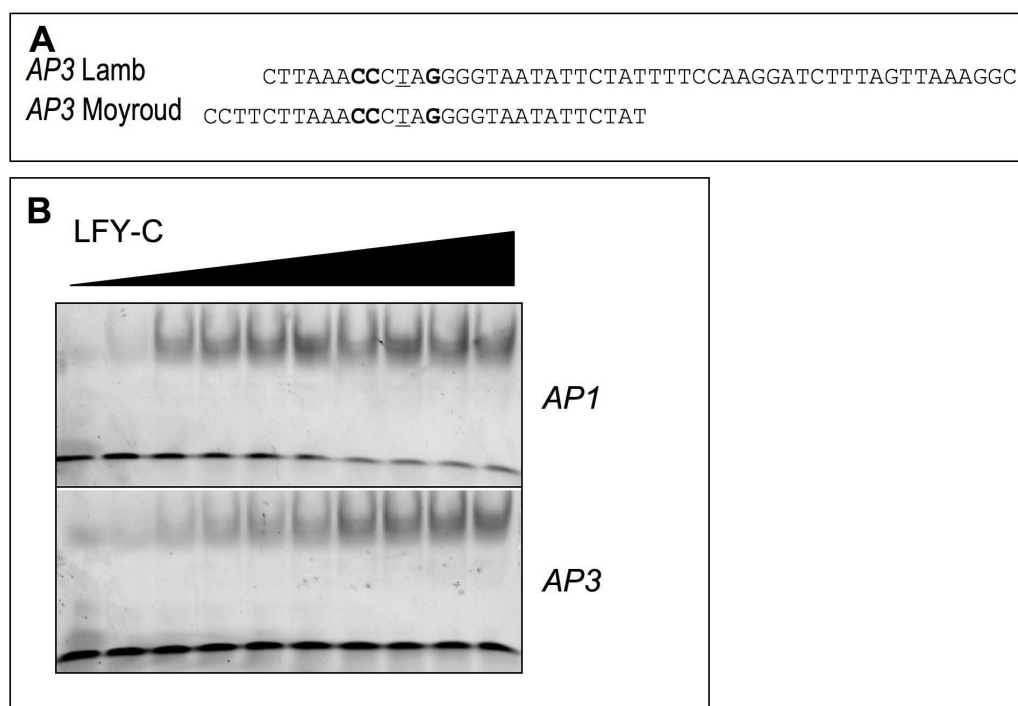


Figure 59: Etude de la liaison de LFY-C à AP3 (travail d'E.Moyroud)

(A) Comparaison de séquence entre les oligonucléotides AP3 dessinés par Lamb et al. (2002) et par E. Moyroud. La séquence médiane départageant les deux demi-sites de liaison de LFY est soulignée, les bases de la séquence consensus de liaison de LFY à l'ADN sont en gras; (B) EMSA réalisé avec 10nM d'oligonucléotides (nature précisée à droite du gel) et une gamme croissante en LFY-C (100, 400, 700, 1000, 2000, 3500, 5000, 6500, 8000nM). La gamme réalisée avec AP1 est représentée ici en référence. AP3 correspond ici à l'oligonucléotide dessiné par E.Moyroud (AP3 Moyroud).

Par ailleurs, Edwige est parvenue à produire et purifier la protéine UFO en bactérie *E. coli*, et tente désormais de visualiser de l'interaction LFY/UFO par EMSA. La cartographie des régions protéiques nécessaires à l'interaction entre LFY et UFO sera donc abordée par approche biochimique couplée à une mutagenèse dirigée de régions ou résidus de LFY.

Nous souhaiterions également valider l'hypothèse de l'ubiquitination de LFY par UFO en combinant une étude biochimique à une analyse de spectrométrie de masse. Si l'ubiquitination de LFY

est confirmée, nous étudierons la fonction de cet étiquetage dans l'activation de LFY. En l'état actuel des connaissances, l'ubiquitination peut stimuler l'activité d'une protéine par deux voies distinctes: soit elle conduit à sa dégradation et ainsi à une accélération de son *turn over* (Kodadek and Bachhawat-Sikder, 2006), soit elle entraîne en elle même l'activation de la protéine (Garcia-Higuera et al., 2001).

c) A la recherche de nouveaux corégulateurs

Il est donc acquis que LFY agit avec les co-régulateurs UFO et WUS dans l'activation de l'expression respectivement des gènes *AP3* et *AG*. Mais qu'en est-il dans le cas d'*AP1*? Des expériences de transactivation dans la levure (où le gène rapporteur codant pour l'activité *lacZ* est sous contrôle d'un promoteur de choix) ont montré que LFY n'était pas capable à lui seul d'activer *AP1*. Il semble donc très probable que dans cette régulation, LFY nécessite également l'intervention d'un corégulateur.

Pour identifier de façon exhaustive les différents corégulateurs de LFY, nous souhaiterions mettre au point un protocole de purification d'une protéine LFY étiquetée produite en plante et analyser les protéines copurifiées. Une autre stratégie pourrait consister à identifier ces corégulateurs par étude bibliographique, en fonction d'indices apportés par les publications passées et futures sur la régulation des différents gènes cibles de LFY.

5. Comment LFY a-t-il évolué ?

La structure de LFY-C a d'ores et déjà permis de mieux comprendre comment LFY d'*A. thaliana* lie l'ADN d'un point de vue moléculaire. Le fait que le domaine C-terminal soit hautement conservé chez les Angiospermes laisse envisager que les données exposées ici sont généralisables à tout ou partie des plantes à fleurs. Si cette conservation reste forte avec des végétaux plus primitifs, certains résidus ont néanmoins été la cible de modification au cours de l'évolution de LFY. Dans le but de tester l'hypothèse avancée par la théorie *Mostly Male* (Frohlich, 2000) prédisant un rôle central de LFY dans l'apparition de la fleur, il est nécessaire d'identifier les résidus ayant pu expliquer l'évolution fonctionnelle de LFY. Jusqu'ici ce travail ne pouvait se faire que par l'étude exhaustive de chacun des résidus ayant évolué entre les différents groupes végétaux sans avoir idée de leur fonction, et sans pouvoir faire d'éventuels liens moléculaires entre des résidus ayant pu coévoluer. Dans ce contexte, la structure constitue un précieux outil à l'identification des résidus variants dont la position laisse présager un rôle important à la fonctionnalité de LFY.

En alliant l'analyse d'un alignement de séquence entre des orthologues de LFY et les données structurales, les résidus à cibler dans une approche évo-dévo peuvent être classés en plusieurs catégories:

i) les résidus impliqués dans la reconnaissance de l'ADN. Ces résidus pourraient être responsables d'une évolution de l'affinité et/ou de la spécificité de liaison de l'ADN ayant pour

conséquence une évolution des gènes cibles de LFY, voire une évolution de sa qualité même de facteur de transcription. Rappelons en effet que dans leur article, Maizel et collaborateurs (2005) n'ont pu mettre en évidence de liaison de l'homologue *PpLFY1* de la Mousse *P. patens* à des éléments cis régulateurs reconnus par LFY d'*A. thaliana*. D'un point de vue général, la face de LFY en regard à l'ADN est particulièrement bien conservée comparativement à la face opposée potentiellement impliquée dans les contacts de LFY avec des partenaires protéiques, nous y reviendrons dans le point ii. Les résidus au potentiel électrostatique positif sont invariants (Fig.60).

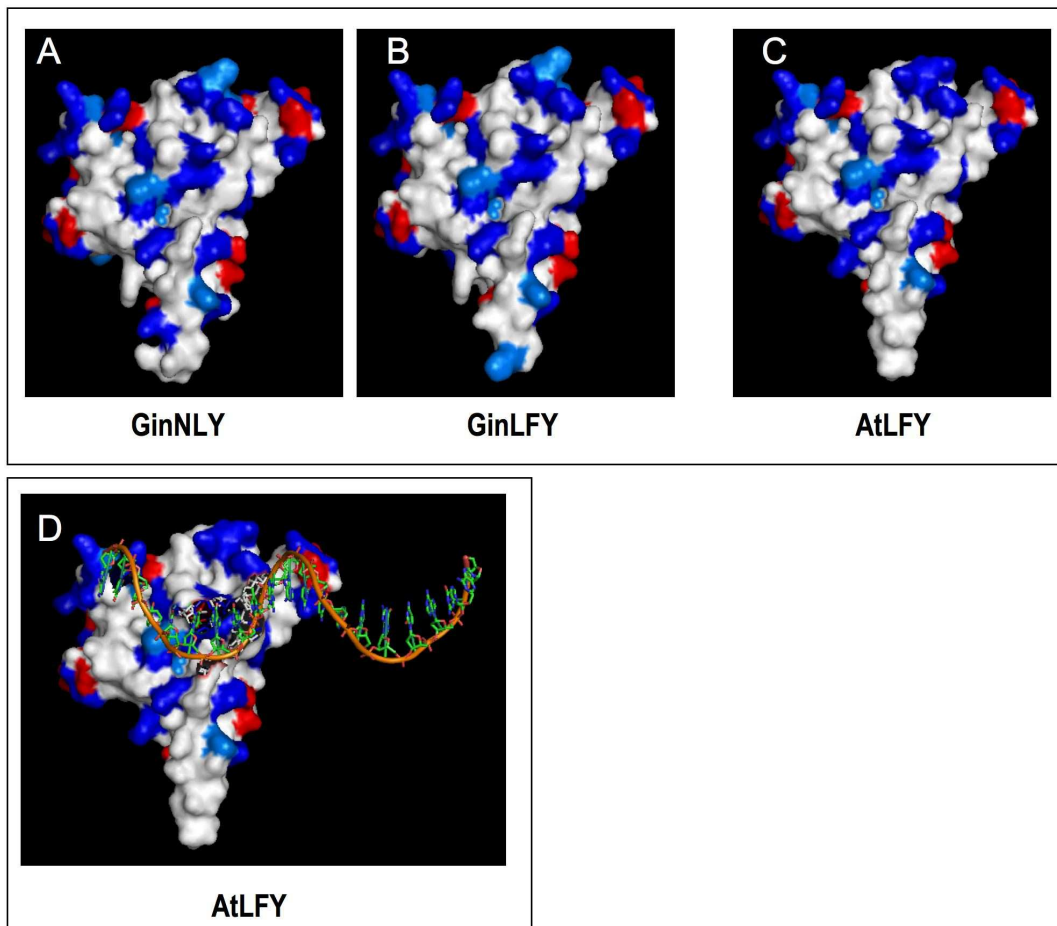


Figure 60: Etude comparative de la face de LFY-C en regard à l'ADN chez *A thaliana* et *G.biloba* (travail de F.Parcy)

Les protéines sont représentées en surface, les lysines et arginines sont représentées en bleu foncé, les histidines en bleu clair, et les aspartates et glutamates en rouge. *G.biloba* est un arbre du groupe des Gymnospermes qui par conséquent possède deux copies du gènes *LFY*, *GinLFY* et *GinNLY*. Leurs représentations sont obtenues par modélisation à partir de la structure de *AtLFY*, à l'aide du logiciel SwissModel. La position de l'ADN (en orange) sur LFY-C est représentée en D. La comparaison entre *AtLFY* et ses homologues chez *G.biloba* révèle une large conservation des résidus chargés positivement, favorisant une interaction avec le double brin d'ADN.

En regardant de plus près l'évolution des résidus impliqués dans l'interaction de LFY à l'ADN, la conservation est là aussi tout à fait saisissante, qu'il s'agisse des résidus impliqués dans des contacts bases spécifiques (Arg237, Asp291, Lys307, Pro308), des résidus impliqués dans des contacts avec le squelette phosphate (Glu238, Phe241, Lys284, Arg295, Asp306, Lys309, Tyr313, Val341, Gly342, Tyr377, Thr380), ou encore des résidus dont le contact avec l'ADN n'est que putatif pour le moment

(Arg235, Lys253 et Lys254). Mentionnons simplement que trois de ces résidus ont été la cible de variations chez les différents sous groupe Angiospermes: la Lys307 est parfois convertie en un autre résidu basique (une arginine), l'Asp291 a subi quelques variations mettant systématiquement en jeu des résidus polaires, et la Pro308 est parfois remplacée par un autre résidu de faible encombrement stérique lui aussi, une sérine. Dans le cas particulier des *Browneas*, la Lys307 qui contacte les deux guanines invariantes ± 2 et ± 3 et la Pro308 qui exerce une pression physique sur les bases ± 5 et ± 6 chez LFY d'*A. thaliana* sont toutes deux remplacées par une Arg307 et une Ser308, respectivement. Il est donc vraisemblable que ces orthologues de LFY de Brownéas reconnaissent des séquences différentes. Il serait intéressant de déterminer leur spécificité et de rechercher si les éléments cis des promoteurs se sont adaptés à d'éventuels changements.

Les points de contacts entre LFY et l'ADN ne semblent pas, *a priori*, avoir été sujet à des modifications majeures. Ce constat ne signifie pas pour autant que les propriétés de liaison de LFY à l'ADN sont restées inchangées depuis l'apparition de LFY. En effet, une modification des résidus impliqués dans la coopérativité de liaison de LFY pourrait se répercuter sur la séquence cible reconnue par LFY. Des orthologues de LFY dont la coopérativité est altérée pourraient ne reconnaître que l'équivalent d'un demi-site d'ADN, ou pourraient nécessiter un espacement différent entre les deux demi-sites. Pour ces raisons, Edwige Moyourd a entrepris une étude complémentaire des propriétés de liaison à l'ADN de différents orthologues de LFY caractéristiques d'un groupe végétal (Angiospermes basaux, Gymnospermes, Mousses) par selex et par approche biochimique classique afin de déterminer respectivement la séquence consensus d'ADN qu'elles reconnaissent et leur comportement de liaison à l'ADN (en considérant les gènes cibles connus de LFY d'*A. thaliana*).

ii) les résidus impliqués dans l'oligomérisation et la coopérativité de liaison de LFY à l'ADN. La coopérativité de liaison de LFY à l'ADN mise en évidence au cours de mon travail de thèse résulte principalement d'une dimérisation de LFY impliquant deux résidus de l'hélice $\alpha 7$, l'His387 et l'Arg390. Nous souhaiterions savoir si cette coopérativité est une acquisition récente de la protéine LFY. Pour cela, nous avons étudié la conservation des deux résidus clés. L'His387 est extrêmement bien conservée depuis les homologues de Mousses. En revanche, l'Arg390 qui est elle aussi très conservée chez les Angiospermes, est remplacée en Lys390 chez de nombreuses Mousses, Fougères et Gymnospermes. Même si les deux chaînes latérales possèdent un caractère basique, leur longueur varie, la lysine comportant un résidu du squelette carboné en moins par rapport à l'arginine. La lysine qui offre par conséquent une plus petite extension de sa chaîne latérale pourrait entraver la capacité de LFY à se dimériser. Cet effet pourrait être accentué par la modification d'un second résidu de l'hélice $\alpha 6$ dont le rôle pourrait être important au bon positionnement de l'hélice $\alpha 7$. L'Asp364 (hélice $\alpha 6$) contacte l'Arg383 (hélice $\alpha 7$) via un pont salin (Fig.61B). Or cet Asp364 est remplacée par un Glu364 chez la

plupart des Fougères et Gymnospermes, résidu conservant la propriété acide de l'aspartate mais présentant cette fois ci un atome de carbone supplémentaire dans sa chaîne latérale par rapport à l'aspartate. Nous postulons qu'une telle modification pourrait éloigner l'hélice $\alpha 7$ du monomère opposé. Il sera alors intéressant d'étudier la coévolution des résidus 364 et 390 chez les différents groupes végétaux (Fig.61A), et d'introduire ces différentes configurations chez la protéine LFY d'*A. thaliana* pour en étudier les conséquences sur le comportement de liaison à l'ADN.

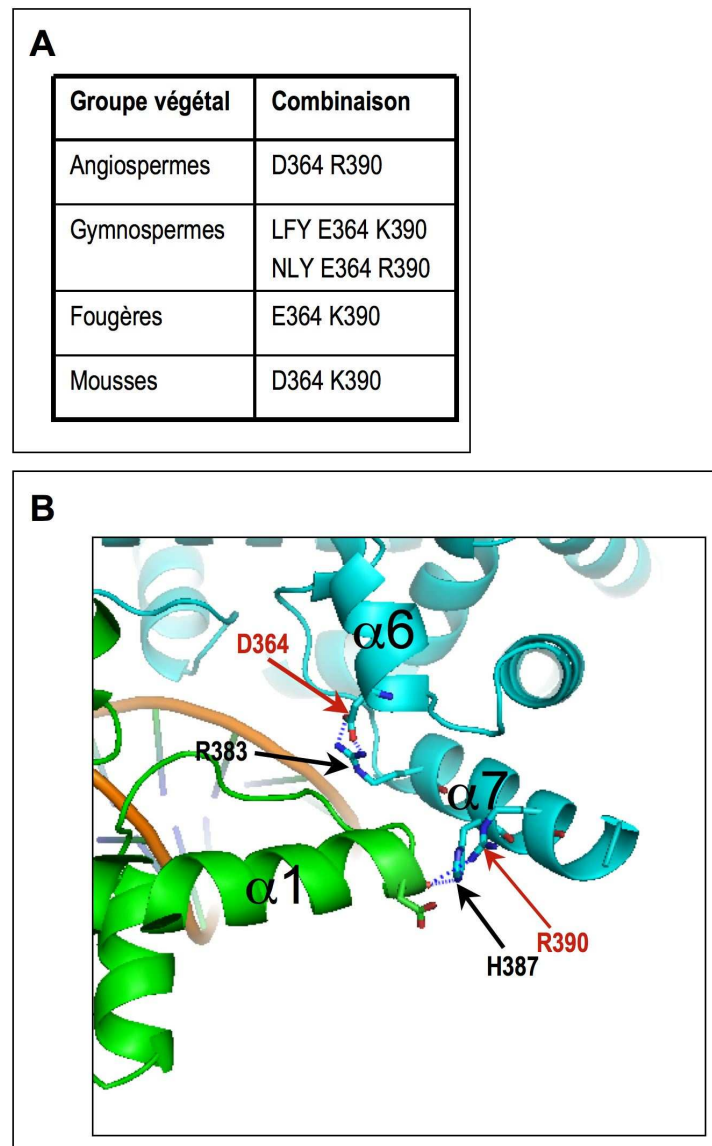


Figure 61: Analyse des résidus 364 et 390 ayant pu jouer un rôle dans l'évolution du comportement coopératif de liaison de LFY à l'ADN

(A) Tableau récapitulatif des combinaisons majoritairement observées en fonction du groupe végétal; (B) Gros plan sur la région d'intérêt obtenu à partir du logiciel PyMOL. Un premier monomère LFY-C est représenté en vert, le second en cyan. Les résidus D364 et R390 d'*A. thaliana* sont indiqués par des flèches rouges, les résidus très conservés R383 et H387 en noir.

Des résultats à venir nous informeront de l'éventuelle capacité de LFY à lier l'ADN selon une coopérativité de liaison mettant en jeu des complexes d'ordre supérieur. Cette éventualité sera également envisagée sous une orientation évo-dévo. Dans leur article, Maizel et collaborateurs font état

de la variation d'une histidine de LFY d'Angiospermes convertie en aspartate chez les Mousses (H312D, correspondant à H394D dans leur article). Ces auteurs montrent que la mutation de cet aspartate en histidine chez l'orthologue de la Mousse *P. patens* entraîne une restauration partielle de la capacité de cette protéine à activer la transcription et une complémentation partielle du phénotype induit par la mutation de *Ify* chez *A. thaliana*. Pour comprendre l'effet de cette mutation, Edwige Moyroud a étudié le comportement de liaison d'une protéine *PpLFY1* de *P. patens* mutée D312H et d'une protéine *AtLFY* d'*A. thaliana* mutée H312D. Je n'évoquerai ici que les résultats obtenus avec la protéine LFY H312D. LFY H312D lie *AP1* selon une affinité comparable à celle d'*AtLFY*. En revanche, à la différence d'*AtLFY* qui à très forte concentration (5000nM) était capable de former un complexe d'ordre supérieur sur l'ADN (Fig.39), la protéine mutée H312D ne parvient jamais à former ce complexe supérieur. Ce résidu His312 chez *A. thaliana* pourrait être impliqué dans une oligomérisation d'ordre supérieur, l'effet de sa mutation sera donc intégré à l'étude du rôle de ce complexe d'ordre supérieur.

iii) les résidus impliqués dans l'interaction de LFY avec des partenaires protéiques. Ces résidus offrent très probablement le plus large champ de prospection, d'une part car aucun résidu impliqué dans des interactions avec les corégulateurs UFO et WUS connus n'a été identifié, et d'autre part car il est probable que ces résidus soient les plus sujets à des variations au cours de l'évolution de LFY. En effet, si les résidus impliqués dans le repliement de LFY, la dimérisation et la liaison à l'ADN sont fortement conservés, les résidus exposés à la surface de LFY ont subi des variations pour près de 70% d'entre eux (estimation réalisée par François Parcy, Fig.62).

Ces variations peuvent s'observer à tous les niveaux, pouvant être spécifiques d'un groupe, d'un genre ou d'une espèce végétale. La grande difficulté ici repose donc sur l'identification des résidus qui effectivement ont joué un rôle dans l'interaction entre LFY et ses partenaires. Quelle stratégie adopter ? Faut-il se concentrer sur les résidus conservés ou au contraire sur les résidus variants ? Cette question est d'autant plus délicate qu'on ne sait si LFY a coévolué avec ses partenaires protéiques ou si au contraire il a changé de partenaires protéiques au cours de l'évolution. Pour simplifier ce travail, le plus raisonnable sera sans doute de commencer dans un premier temps par identifier les résidus impliqués dans les contacts éventuels de LFY avec UFO et/ou WUS et de se concentrer sur leur rôle au cours de l'évolution de LFY.

iv) les résidus impliqués dans le repliement de LFY. Les résidus faisant l'architecture de LFY sont extrêmement conservés. Néanmoins, les deux résidus variants H312D et C345R chez la Mousse *P. patens* identifiés par Maizel et collaborateurs (2005) pour intervenir dans le bon repliement du domaine C-terminal de LFY. En effet, nous avons constaté que ces résidus interagissaient l'un avec l'autre. Etant donné que l'His312 fait partie de l'hélice $\alpha 3$ établissant d'importants contacts avec l'ADN, perturber l'interaction entre ce résidu et la Cys345 pourrait occasionner une modification du repliement

du domaine C-terminal et ainsi expliquer une évolution des propriétés de liaison entre AtLFY et PpLFY1.

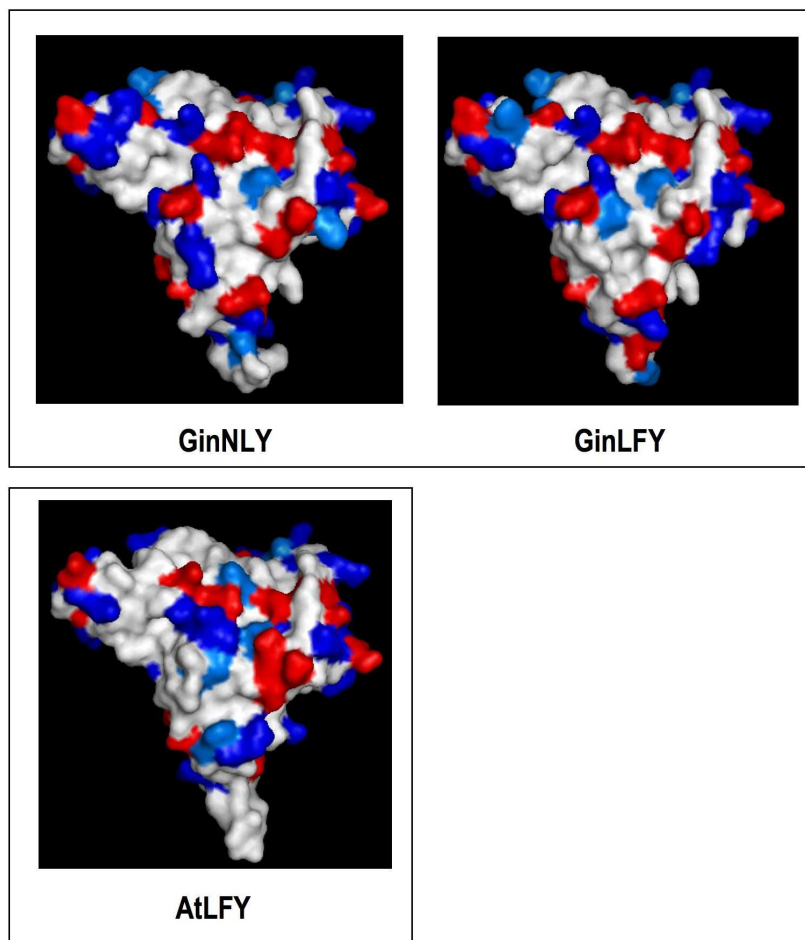


Figure 62: Etude comparative de la face extérieure de LFY-C (opposée à l'ADN) chez *A.thaliana* et *G.biloba* (travail de F.Parcy)

Les protéines sont représentées en surface, les lysines et arginines sont représentées en bleu foncé, les histidines en bleu clair, et les aspartates et glutamates en rouge. *G.biloba* est un arbre du groupe des Gymnospermes qui par conséquent possède deux copies du gènes *LFY*, *GinLFY* et *GinNLY*. Leurs représentations sont obtenues par modélisation à partir de la structure de AtLFY, à l'aide du logiciel SwissModel. La comparaison entre AtLFY et ses homologues chez *G.biloba* révèle d'importants changements dans la répartition des résidus chargés.

6. Quelle relation existe-il entre LFY et l'apparition des plantes à fleurs ?

A mon avis, les Angiospermes ou plantes à fleurs constituent à n'en pas douter la plus belle invention de la nature ! L'apparition de la fleur constitue un avantage sélectif sans égal, succès à l'origine d'une véritable invasion des Angiospermes qui en l'espace de 120 millions d'années ont développé pas moins de 250000 espèces. Selon l'arbre phylogénétique des plantes terrestres, les plantes à fleur sont monophylétiques et ne seraient donc apparues qu'au niveau d'un seul foyer de localisation inconnue. C'est cette origine unique qui expliquerait le fossé morphologique existant aujourd'hui entre les Angiospermes et leurs prédécesseurs, les Gymnospermes. Malgré les efforts entrepris par les botanistes et biologistes-théoriciens de l'évolution, les mécanismes mis en place par

les plantes pour expliquer une telle apparition n'ont pas été élucidés. Cette recherche est d'autant plus difficile qu'on ne connaît que très peu de choses sur l'ancêtre des Angiospermes et l'ancêtre commun des Angiospermes et Gymnospermes et que les fossiles connus apportent peu de renseignements quant à leurs caractéristiques morphologiques. Le seul recours à la compréhension de l'apparition des fleurs sur terre semble donc une étude fonctionnelle évolutive des acteurs importants à cette transition.

Les gènes ABCE responsables de l'identité des organes floraux sont une cible de choix, et LFY en tant que régulateur de l'expression de ces gènes en est une autre.

Mon travail de thèse apporte des éléments de réflexion utiles à la compréhension des mécanismes ayant conduit à l'évolution des fonctions de LFY. Cependant expliquer comment LFY a évolué exigera un investissement de plusieurs années, voire décennies. En plus de l'étude sur le domaine conservé C-terminal, il nécessitera l'étude de régions nettement moins caractérisées comme la région conservée N-terminale et les régions non conservées. La difficulté sera d'autant plus accrue que cette évolution peut impliquer la coévolution de résidus qui indépendamment n'ont pas apporté de transition fonctionnelle majeure.

Pour soulager en partie ces difficultés, l'approche de synthèse fonctionnelle (*'functional synthesis'*; Dean and Thornton, 2007) alimente de grands espoirs sur le principe de résurrection de protéines éteintes. Il est pour cela nécessaire de prédire à partir d'un alignement de séquences entre homologues de LFY issus des principaux groupes végétaux quelle était la séquence des protéines LFY présentent chez l'ancêtre des Angiospermes et l'ancêtre commun des Angiospermes et Gymnospermes. Parvenir à produire ces protéines 'ancestrales' et à analyser leurs propriétés par approche biochimique pourrait apporter des éléments de compréhension sur l'évolution du mode de fonctionnement de LFY, ce qui constituerait de précieuses indications sur le rôle évolutif joué par cette protéine. L'"abominable mystère" de Charles Darwin trouvera peut-être un jour une solution !

Références bibliographiques

- Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana* (2000). The *Arabidopsis* Genome Initiative. *Nature*, **408**, 796-815.
- Abe, M., Kobayashi, Y., Yamamoto, S., Daimon, Y., Yamaguchi, A., Ikeda, Y., Ichinoki, H., Notaguchi, M., Goto, K. and Araki, T. (2005) FD, a bZIP protein mediating signals from the floral pathway integrator FT at the shoot apex. *Science*, **309**, 1052-1056.
- Ahn, J.H., Miller, D., Winter, V.J., Banfield, M.J., Lee, J.H., Yoo, S.Y., Henz, S.R., Brady, R.L. and Weigel, D. (2006) A divergent external loop confers antagonistic activity on floral regulators FT and TFL1. *Embo J*, **25**, 605-614.
- Albert, B., Godelle, B. and Gouyon, P. (1998) Evolution of the plant mitochondrial genome: dynamics of duplication and deletion of sequences. *J Mol Evol*, **46**, 155-158.
- Alvarez-Buylla, E.R., Garcia-Ponce, B. and Garay-Arroyo, A. (2006) Unique and redundant functional domains of APETALA1 and CAULIFLOWER, two recently duplicated *Arabidopsis thaliana* floral MADS-box genes. *J Exp Bot*, **57**, 3099-3107.
- Amasino, R. (2004a) Take a cold flower. *Nat Genet*, **36**, 111-112.
- Amasino, R. (2004b) Vernalization, competence, and the epigenetic memory of winter. *Plant Cell*, **16**, 2553-2559.
- Ansaldi, M., Simon, G., Lepelletier, M. and Mejean, V. (2000) The TorR high-affinity binding site plays a key role in both *torR* autoregulation and *torCAD* operon expression in *Escherichia coli*. *J Bacteriol*, **182**, 961-966.
- Aravind, L., Anantharaman, V., Balaji, S., Babu, M.M. and Iyer, L.M. (2005) The many faces of the helix-turn-helix domain: transcription regulation and beyond. *FEMS Microbiol. Rev.*, **29**, 231-262.
- Ausin, I., Alonso-Blanco, C. and Martinez-Zapater, J.M. (2005) Environmental regulation of flowering. *Int J Dev Biol*, **49**, 689-705.
- Baker, M.D., Gendlina, I., Collins, C.M. and Acharya, K.R. (2004) Crystal structure of a dimeric form of streptococcal pyrogenic exotoxin A (SpeA1). *Protein Sci*, **13**, 2285-2290.
- Baum, D.A. and Hileman, L.C. (2006) Flowering and its Manipulation. Blackwell, Oxford.
- Ben-Naim, O., Eshed, R., Parnis, A., Teper-Bamnolker, P., Shalit, A., Coupland, G., Samach, A. and Lifschitz, E. (2006) The CCAAT binding factor can mediate interactions between CONSTANS-like proteins and DNA. *Plant J*, **46**, 462-476.
- Bernier, G. and Périlleux, C. (2005) A physiological overview of the genetics of flowering time control. *Plant Biotechnology Journal*, **3**, 3-16.
- Blazquez, M.A., Ahn, J.H. and Weigel, D. (2003) A thermosensory pathway controlling flowering time in *Arabidopsis thaliana*. *Nat Genet*, **33**, 168-171.
- Blazquez, M.A., Green, R., Nilsson, O., Sussman, M.R. and Weigel, D. (1998) Gibberellins promote flowering of *Arabidopsis* by activating the *LEAFY* promoter. *Plant Cell*, **10**, 791-800.
- Blazquez, M.A., Soowal, L.N., Lee, I. and Weigel, D. (1997) *LEAFY* expression and flower initiation in *Arabidopsis*. *Development*, **124**, 3835-3844.
- Blazquez, M.A. and Weigel, D. (2000) Integration of floral inductive signals in *Arabidopsis*. *Nature*, **404**, 889-892.
- Bomblies, K., Wang, R.L., Ambrose, B.A., Schmidt, R.J., Meeley, R.B. and Doebley, J. (2003) Duplicate *FLORICAULA/LEAFY* homologs *zfl1* and *zfl2* control inflorescence architecture and flower patterning in maize. *Development*, **130**, 2385-2395.
- Borner, R., Kampmann, G., Chandler, J., Gleissner, R., Wisman, E., Apel, K. and Melzer, S. (2000) A MADS domain gene involved in the transition to flowering in *Arabidopsis*. *Plant J*, **24**, 591-599.
- Boss, P.K., Bastow, R.M., Mylne, J.S. and Dean, C. (2004) Multiple pathways in the decision to flower: enabling, promoting, and resetting. *Plant Cell*, **16 Suppl**, S18-31.
- Bourne, Y., Watson, M.H., Hickey, M.J., Holmes, W., Rocque, W., Reed, S.I. and Tainer, J.A. (1996) Crystal structure and mutational analysis of the human CDK2 kinase complex with cell cycle-regulatory protein CksHs1. *Cell*, **84**, 863-874.
- Bowman, J.L., Drews, G.N. and Meyerowitz, E.M. (1991) Expression of the *Arabidopsis* floral homeotic gene *AGAMOUS* is restricted to specific cell types late in flower development. *Plant Cell*, **3**, 749-758.
- Bowman, J.L. and Meyerowitz, E.M. (1991) Genetic control of pattern formation during flower development in *Arabidopsis*. *Symp Soc Exp Biol*, **45**, 89-115.
- Bowman, J.L., Smyth, D.R. and Meyerowitz, E.M. (1989) Genes directing flower development in *Arabidopsis*. *Plant Cell*, **1**, 37-52.
- Bradford, M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry*, **72**, 248-254.
- Brown, A.K., Meng, G., Ghadbane, H., Scott, D.J., Dover, L.G., Nigou, J., Besra, G.S. and Futterer, K. (2007) Dimerization of inositol monophosphatase *Mycobacterium tuberculosis* SuhB is not constitutive, but induced by binding of the activator Mg²⁺. *BMC Struct Biol*, **7**, 55.
- Burz, D.S., Rivera-Pomar, R., Jackle, H. and Hanes, S.D. (1998) Cooperative DNA-binding by Bicoid provides a mechanism for threshold-dependent gene activation in the *Drosophila* embryo. *Embo J*, **17**, 5998-6009.
- Busch, M.A., Bomblies, K. and Weigel, D. (1999) Activation of a floral homeotic gene in *Arabidopsis*. *Science*, **285**, 585-587.
- Campbell, R.E., Tour, O., Palmer, A.E., Steinbach, P.A., Baird, G.S., Zacharias, D.A. and Tsien, R.Y. (2002) A monomeric red fluorescent protein. *PNAS*, **99**, 7877-7882.
- Canet, D., Doering, K., Dobson, C.M. and Dupont, Y. (2001) High-sensitivity fluorescence anisotropy detection of protein-folding events: application to alpha-lactalbumin. *Biophys J*, **80**, 1996-2003.
- Carroll, S.B. (2000) Endless forms: the evolution of gene regulation and morphological diversity. *Cell*, **101**, 577-580.
- Chae, E., Tan, Q.K., Hill, T.A. and Irish, V.F. (2008) An *Arabidopsis* F-box protein acts as a transcriptional co-factor to regulate floral development. *Development*, **135**, 1235-1245.
- Cherry, J.L. and Adler, F.R. (2000) How to make a biological switch. *J. Theor. Biol.*, **203**, 117-133.
- Chujo, A., Zhang, Z., Kishino, H., Shimamoto, K. and Kyoizuka, J. (2003) Partial conservation of *LFY* function between rice and *Arabidopsis*. *Plant Cell Physiol*, **44**, 1311-1319.
- Clarke, J.H. and Dean, C. (1994) Mapping *FRI*, a locus controlling flowering time and vernalization response

- in *Arabidopsis thaliana*. *Mol. Gen. Genet.*, **242**, 81-89.
- Clarke N. D., Kissinger C. R., Desjarlais J., Gilliland G. L. and O., P.C. (1994) Structural studies of the engrailed homeodomain. *Protein Sci*, **3**, 1779-1787.
- Coen, E.S. and Meyerowitz, E.M. (1991) The war of the whorls: genetic interactions controlling flower development. *Nature*, **353**, 31-37.
- Coen, E.S., Romero, J.M., Doyle, S., Elliott, R., Murphy, G. and Carpenter, R. (1990) *FLORICAULA*: a homeotic gene required for flower development in *Antirrhinum majus*. *Cell*, **63**, 1311-1322.
- Corbesier, L., Vincent, C., Jang, S., Fornara, F., Fan, Q., Searle, I., Giakountis, A., Farrona, S., Gissot, L., Turnbull, C. and Coupland, G. (2007) FT protein movement contributes to long-distance signaling in floral induction of *Arabidopsis*. *Science*, **316**, 1030-1033.
- Cranz, S., Berger, C., Baici, A., Jelesarov, I. and Bosshard, H.R. (2004) Monomeric and dimeric bZIP transcription factor GCN4 bind at the same rate to their target DNA site. *Biochemistry*, **43**, 718-727.
- Dean, A.M. and Thornton, J.W. (2007) Mechanistic approaches to the study of evolution: the functional synthesis. *Nat. Rev. Genet.*, **8**, 675-688.
- Deshaies, R.J. (1999) SCF and Cullin/Ring H2-based ubiquitin ligases. *Annu Rev Cell Dev Biol*, **15**, 435-467.
- Ditta, G., Pinyopich, A., Robles, P., Pelaz, S. and Yanofsky, M.F. (2004) The *SEP4* gene of *Arabidopsis thaliana* functions in floral organ and meristem identity. *Curr Biol*, **14**, 1935-1940.
- Dong, Z.C., Zhao, Z., Liu, C.W., Luo, J.H., Yang, J., Huang, W.H., Hu, X.H., Wang, T.L. and Luo, D. (2005) Floral patterning in *Lotus japonicus*. *Plant Physiol*, **137**, 1272-1282.
- Dornelas, M.C. and Rodriguez, A.P. (2005) The rubber tree (*Hevea brasiliensis* Muell. Arg.) homologue of the *LEAFY/FLORICAULA* gene is preferentially expressed in both male and female floral meristems. *J Exp Bot*, **56**, 1965-1974.
- Dornelas, M.C. and Rodriguez, A.P. (2006) The tropical cedar tree (*Cedrela fissilis* Vell., *Meliaceae*) homolog of the *Arabidopsis* *LEAFY* gene is expressed in reproductive tissues and can complement *Arabidopsis* leafy mutants. *Planta*, **223**, 306-314.
- Dummler, A., Lawrence, A.M. and de Marco, A. (2005) Simplified screening for the detection of soluble fusion constructs expressed in *E. coli* using a modular set of vectors. *Microb Cell Fact*, **4**, 34.
- Eriksson, S., Bohlenius, H., Moritz, T. and Nilsson, O. (2006) GA4 is the active gibberellin in the regulation of *LEAFY* transcription and *Arabidopsis* floral initiation. *Plant Cell*, **18**, 2172-2181.
- Feild, T.S., Brodribb, T., and Holbrook, M. (2002) Hardly a relict: freezing and the evolution of vesselless wood in Winteraceae. *Evolution*, **56**, 464-478.
- Ferrandiz, C., Gu, Q., Martienssen, R. and Yanofsky, M.F. (2000) Redundant regulation of meristem identity and plant architecture by *FRUITFULL*, *APETALA1* and *CAULIFLOWER*. *Development*, **127**, 725-734.
- Fraenkel, E., Rould, M.A., Chambers, K.A. and Pabo, C.O. (1998) Engrailed homeodomain-DNA complex at 2.2 Å resolution: a detailed view of the interface and comparison with other engrailed structures. *J Mol Biol*, **284**, 351-361.
- Frohlich, M.W., and Parker, D.S. (2000) The Mostly male theory of flower evolutionary origins: from genes to fossils. *Systematic Botany*, **25**, 155-170.
- Frohlich, M.W. and Chase, M.W. (2007) After a dozen years of progress the origin of angiosperms is still a great mystery. *Nature*, **450**, 1184-1189.
- Garcia-Higuera, I., Taniguchi, T., Ganesan, S., Meyn, M.S., Timmers, C., Hejna, J., Grompe, M. and D'Andrea, A.D. (2001) Interaction of the *Fanconi anemia* proteins and BRCA1 in a common pathway. *Mol Cell*, **7**, 249-262.
- Garman, E.F. and Grime, G.W. (2005) Elemental analysis of proteins by microPIXE. *Prog Biophys Mol Biol*, **89**, 173-205.
- Gehring, W.J., Qian, Y.Q., Billeter, M., Furukubo-Tokunaga, K., Schier, A.F., Resendez-Perez, D., Affolter, M., Otting, G. and Wuthrich, K. (1994) Homeodomain-DNA recognition. *Cell*, **78**, 211-223.
- Gerard, F.C., Ribeiro Ede, A., Jr., Albertini, A.A., Gutsche, I., Zaccai, G., Ruigrok, R.W. and Jamin, M. (2007) Unphosphorylated rhabdoviridae phosphoproteins form elongated dimers in solution. *Biochemistry*, **46**, 10328-10338.
- Gietz, R.D., Schiestl, R.H., Willems, A.R. and Woods, R.A. (1995) Studies on the transformation of intact yeast cells by the LiAc/SS-DNA/PEG procedure. *Yeast*, **11**, 355-360.
- Gocal, G.F., Sheldon, C.C., Gubler, F., Moritz, T., Bagnall, D.J., MacMillan, C.P., Li, S.F., Parish, R.W., Dennis, E.S., Weigel, D. and King, R.W. (2001) *GAMYB*-like genes, flowering, and gibberellin signaling in *Arabidopsis*. *Plant Physiol*, **127**, 1682-1693.
- Gomez-Mena, C., Pineiro, M., Franco-Zorrilla, J.M., Salinas, J., Coupland, G. and Martinez-Zapater, J.M. (2001) *early bolting in short days*: an *Arabidopsis* mutation that causes early flowering and partially suppresses the floral phenotype of leafy. *Plant Cell*, **13**, 1011-1024.
- Gouaux, E. (2004) Structure and function of AMPA receptors. *J Physiol*, **554**, 249-253.
- Grant, V. (1994) Modes and origins of mechanical and ethological isolation in angiosperms. *PNAS*, **91**, 3-10.
- Gregory, P.D., Barbaric, S. and Horz, W. (1998) Analyzing chromatin structure and transcription factor binding in yeast. *Methods*, **15**, 295-302.
- Gregory, P.D., Wagner, K. and Horz, W. (2001) Histone acetylation and chromatin remodeling. *Exp Cell Res*, **265**, 195-202.
- Gromiha, M.M., Siebers, J.G., Selvaraj, S., Kono, H. and Sarai, A. (2005) Role of inter and intramolecular interactions in protein-DNA recognition. *Gene*, **364**, 108-113.
- Gu, Q., Ferrandiz, C., Yanofsky, M.F. and Martienssen, R. (1998) The *FRUITFULL* MADS-box gene mediates cell differentiation during *Arabidopsis* fruit development. *Development*, **125**, 1509-1517.
- Guarente, L. (1983) Yeast promoters and *lacZ* fusions designed to study expression of cloned genes in yeast. *Methods Enzymol*, **101**, 181-191.
- Hartmann, U., Hohmann, S., Nettesheim, K., Wisman, E., Saedler, H. and Huijser, P. (2000) Molecular cloning of *SVP*: a negative regulator of the floral transition in *Arabidopsis*. *Plant J*, **21**, 351-360.
- He, Y., Michaels, S.D. and Amasino, R.M. (2003) Regulation of flowering time by histone acetylation in *Arabidopsis*. *Science*, **302**, 1751-1754.
- Helliwell, C.A., Wood, C.C., Robertson, M., James Peacock, W. and Dennis, E.S. (2006) The *Arabidopsis* FLC protein interacts directly *in vivo* with *SOC1* and *FT* chromatin and is part of a high-molecular-weight protein complex. *Plant J*, **46**, 183-192.

- Henderson, I.R. and Dean, C. (2004) Control of *Arabidopsis* flowering: the chill before the bloom. *Development*, **131**, 3829-3838.
- Hepworth, S.R., Klenz, J.E. and Haughn, G.W. (2006) *UFO* in the *Arabidopsis* inflorescence apex is required for floral-meristem identity and bract suppression. *Planta*, **223**, 769-778.
- Hepworth, S.R., Valverde, F., Ravenscroft, D., Mouradov, A. and Coupland, G. (2002) Antagonistic regulation of flowering-time gene *SOC1* by *CONSTANS* and *FLC* via separate promoter motifs. *Embo J*, **21**, 4327-4337.
- Hill, T.A., Day, C.D., Zondlo, S.C., Thackeray, A.G. and Irish, V.F. (1998) Discrete spatial and temporal cis-acting elements regulate transcription of the *Arabidopsis* floral homeotic gene *APETALA3*. *Development*, **125**, 1711-1721.
- Himi, S., Sano, R., Nishiyama, T., Tanahashi, T., Kato, M., Ueda, K. and Hasebe, M. (2001) Evolution of MADS-box gene induction by *FLO/LFY* genes. *J Mol Evol*, **53**, 387-393.
- Hofer, J., Turner, L., Hellens, R., Ambrose, M., Matthews, P., Michael, A. and Ellis, N. (1997) *UNIFOLIATA* regulates leaf and flower morphogenesis in pea. *Curr Biol*, **7**, 581-587.
- Hong, R.L., Hamaguchi, L., Busch, M.A. and Weigel, D. (2003) Regulatory elements of the floral homeotic gene *AGAMOUS* identified by phylogenetic footprinting and shadowing. *Plant Cell*, **15**, 1296-1309.
- Honma, T. and Goto, K. (2000) The *Arabidopsis* floral homeotic gene *PISTILLATA* is regulated by discrete cis-elements responsive to induction and maintenance signals. *Development*, **127**, 2021-2030.
- Honma, T. and Goto, K. (2001) Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. *Nature*, **409**, 525-529.
- Hu, C.D., Chinenov, Y. and Kerppola, T.K. (2002) Visualization of interactions among bZIP and Rel family proteins in living cells using bimolecular fluorescence complementation. *Mol Cell*, **9**, 789-798.
- Huala, E. and Sussex, I.M. (1992) *LEAFY* interacts with Floral Homeotic Genes to Regulate *Arabidopsis* Floral Development. *Plant Cell*, **4**, 901-903.
- Hurst, H.C. (1995) Transcription factors 1: bZIP proteins. *Protein Profile*, **2**, 101-168.
- Ingram, G.C., Goodrich, J., Wilkinson, M.D., Simon, R., Haughn, G.W. and Coen, E.S. (1995) Parallels between *UNUSUAL FLORAL ORGANS* and *FIMBRIATA*, genes controlling flower development in *Arabidopsis* and *Antirrhinum*. *Plant Cell*, **7**, 1501-1510.
- Irish, V.F. and Sussex, I.M. (1990) Function of the *apetala-1* Gene during *Arabidopsis* Floral Development. *Plant Cell*, **2**, 741-753.
- Jacobs, G.H. (1992) Determination of the base recognition positions of zinc fingers from sequence analysis. *Embo J*, **11**, 4507-4517.
- Jaeger, K.E. and Wigge, P.A. (2007) FT protein acts as a long-range signal in *Arabidopsis*. *Curr Biol*, **17**, 1050-1054.
- Kardailsky, I., Shukla, V.K., Ahn, J.H., Dagenais, N., Christensen, S.K., Nguyen, J.T., Chory, J., Harrison, M.J. and Weigel, D. (1999) Activation tagging of the floral inducer *FT*. *Science*, **286**, 1962-1965.
- Kempin, S.A., Mandel, M.A. and Yanofsky, M.F. (1993) Conversion of perianth into reproductive organs by ectopic expression of the tobacco floral homeotic gene *NAG1*. *Plant Physiol.*, **103**, 1041-1046.
- Kempin, S.A., Savidge, B. and Yanofsky, M.F. (1995) Molecular basis of the *cauliflower* phenotype in *Arabidopsis*. *Science*, **267**, 522-525.
- Kirsch, R.D. and Joly, E. (1998) An improved PCR-mutagenesis strategy for two-site mutagenesis or sequence swapping between related genes. *Nucleic Acids Res.*, **26**, 1848-1850.
- Kobayashi, Y., Kaya, H., Goto, K., Iwabuchi, M. and Araki, T. (1999) A pair of related genes with antagonistic roles in mediating flowering signals. *Science*, **286**, 1960-1962.
- Kobayashi, Y. and Weigel, D. (2007) Move on up, it's time for change--mobile signals controlling photoperiod-dependent flowering. *Genes Dev*, **21**, 2371-2384.
- Kodadek, T. and Bachhawat-Sikder, K. (2006) Optimized protocols for the isolation of specific protein-binding peptides or peptoids from combinatorial libraries displayed on beads. *Mol Biosyst*, **2**, 25-35.
- Kohler, J.J., Metallo, S.J., Schneider, T.L. and Schepartz, A. (1999) DNA specificity enhanced by sequential binding of protein monomers. *PNAS*, **96**, 11735-11739.
- Kohler, J.J. and Schepartz, A. (2001) Kinetic studies of Fos-Jun.DNA complex formation: DNA binding prior to dimerization. *Biochemistry*, **40**, 130-142.
- Koncz, C. and Schell, J. (1986) The promoter TL -DNA gene 5 controls the tissue-specific expression of chimaeric genes carried by a novel type of *Agrobacterium* vector. *Mol Gen Genet* **204**.
- Koomneef, M., Hanhart, C.J. and van der Veen, J.H. (1991) A genetic and physiological analysis of late flowering mutants in *Arabidopsis thaliana*. *Mol. Gen. Genet.*, **229**, 57-66.
- Krizek, B.A. and Fletcher, J.C. (2005) Molecular mechanisms of flower development: an armchair guide. *Nat Rev Genet*, **6**, 688-698.
- Lamb, R.S., Hill, T.A., Tan, Q.K. and Irish, V.F. (2002) Regulation of *APETALA3* floral homeotic gene expression by meristem identity genes. *Development*, **129**, 2079-2086.
- Lamber, E.P., Wilmanns, M. and Svergun, D.I. (2008) Low resolution structural models of the basic helix-loop-helix leucine zipper domain of upstream stimulatory factor 1 and its complexes with DNA from small angle X-ray scattering data. *Biophys J*, **94**, 193-197.
- Laufs, P., Coen, E., Kronenberger, J., Traas, J. and Doonan, J. (2003) Separable roles of *UFO* during floral development revealed by conditional restoration of gene function. *Development*, **130**, 785-796.
- Laux, T., Mayer, K.F.X., Berger, J. and Jürgens, G. (1996) The *WUSCHEL* gene is required for shoot and floral meristem integrity in *Arabidopsis*. *Development*, **122**, 87-96.
- Lavy, M., Bracha-Drori, K., Sternberg, H. and Yalovsky, S. (2002) A cell-specific, prenylation-independent mechanism regulates targeting of type II RACs. *Plant Cell*, **14**, 2431-2450.
- Lee, H., Suh, S.S., Park, E., Cho, E., Ahn, J.H., Kim, S.G., Lee, J.S., Kwon, Y.M. and Lee, I. (2000) The AGAMOUS-LIKE 20 MADS domain protein integrates floral inductive pathways in *Arabidopsis*. *Genes Dev*, **14**, 2366-2376.
- Lee, I., Wolfe, D.S., Nilsson, O. and Weigel, D. (1997a) A *LEAFY* co-regulator encoded by *UNUSUAL FLORAL ORGANS*. *Curr Biol*, **7**, 95-104.
- Lee, J.H., Yoo, S.J., Park, S.H., Hwang, I., Lee, J.S. and Ahn, J.H. (2007) Role of SVF in the control of flowering time by ambient temperature in *Arabidopsis*. *Genes Dev*, **21**, 397-402.

- Lenhard, M., Bohnert, A., Jurgens, G. and Laux, T. (2001) Termination of stem cell maintenance in *Arabidopsis* floral meristems by interactions between *WUSCHEL* and *AGAMOUS*. *Cell*, **105**, 805-814.
- Leonard, T.A., Butler, P.J. and Lowe, J. (2004) Structural analysis of the chromosome segregation protein Spo0J from *Thermus thermophilus*. *Mol Microbiol*, **53**, 419-432.
- Levin, J.Z. and Meyerowitz, E.M. (1995) *UFO*: an *Arabidopsis* gene involved in both floral meristem and floral organ development. *Plant Cell*, **7**, 529-548.
- Levy, Y.Y., Mesnage, S., Mylne, J.S., Gendall, A.R. and Dean, C. (2002) Multiple roles of *Arabidopsis* *VRN1* in vernalization and flowering time control. *Science*, **297**, 243-246.
- Liljegren, S.J., Gustafson-Brown, C., Pinyopich, A., Ditta, G.S. and Yanofsky, M.F. (1999) Interactions among *APETALA1*, *LEAFY*, and *TERMINAL FLOWER1* specify meristem fate. *Plant Cell*, **11**, 1007-1018.
- Lin, M.K., Belanger, H., Lee, Y.J., Varkonyi-Gasic, E., Taoka, K., Miura, E., Xoconostle-Cazares, B., Gendler, K., Jorgensen, R.A., Phinney, B., Lough, T.J. and Lucas, W.J. (2007) FLOWERING LOCUS T protein may act as the long-distance florigenic signal in the cucurbits. *Plant Cell*, **19**, 1488-1506.
- Liu, C., Chen, H., Er, H.L., Soo, H.M., Kumar, P.P., Han, J.H., Liou, Y.C. and Yu, H. (2008) Direct interaction of *AGL24* and *SOC1* integrates flowering signals in *Arabidopsis*. *Development*, **135**, 1481-1491.
- Liu, C., Zhou, J., Bracha-Drori, K., Yalovsky, S., Ito, T. and Yu, H. (2007) Specification of *Arabidopsis* floral meristem identity by repression of flowering time genes. *Development*, **134**, 1901-1910.
- Lohmann, J.U., Hong, R.L., Hobe, M., Busch, M.A., Parcy, F., Simon, R. and Weigel, D. (2001) A molecular link between stem cell regulation and floral patterning in *Arabidopsis*. *Cell*, **105**, 793-803.
- Lundblad, J.R., Lurance, M. and Goodman, R.H. (1996) Fluorescence polarization analysis of protein-DNA and protein-protein interactions. *Mol Endocrinol*, **10**, 607-612.
- Luscombe, N.M., Austin, S.E., Berman, H.M. and Thornton, J.M. (2000) An overview of the structures of protein-DNA complexes. *Genome Biol.*, **1**, REVIEWS001.
- Maizel, A., Busch, M.A., Tanahashi, T., Perkovic, J., Kato, M., Hasebe, M. and Weigel, D. (2005) The floral regulator *LEAFY* evolves by substitutions in the DNA binding domain. *Science*, **308**, 260-263.
- Mandel, M.A., Gustafson-Brown, C., Savidge, B. and Yanofsky, M.F. (1992) Molecular characterization of the *Arabidopsis* floral homeotic gene *APETALA1*. *Nature*, **360**, 273-277.
- Mathieu, J., Warthmann, N., Kuttner, F. and Schmid, M. (2007) Export of FT protein from phloem companion cells is sufficient for floral induction in *Arabidopsis*. *Curr Biol*, **17**, 1055-1060.
- Mayer, K.F., Schoof, H., Haecker, A., Lenhard, M., Jurgens, G. and Laux, T. (1998) Role of *WUSCHEL* in regulating stem cell fate in the *Arabidopsis* shoot meristem. *Cell*, **95**, 805-815.
- McGonigle, B., Bouhidel, K. and Irish, V.F. (1996) Nuclear localization of the *Arabidopsis* *APETALA3* and *PISTILLATA* homeotic gene products depends on their simultaneous expression. *Genes Dev.*, **10**, 1812-1821.
- Mellerowicz, E.J., Horgan, K., Walden, A., Coker, A. and Walter, C. (1998) *PRFLL*--a *Pinus radiata* homologue of *FLORICAULA* and *LEAFY* is expressed in buds containing vegetative shoot and undifferentiated male cone primordia. *Planta*, **206**, 619-629.
- Michaels, S.D. and Amasino, R.M. (1999) FLOWERING LOCUS C encodes a novel MADS domain protein that acts as a repressor of flowering. *Plant Cell*, **11**, 949-956.
- Michaels, S.D., Ditta, G., Gustafson-Brown, C., Pelaz, S., Yanofsky, M. and Amasino, R.M. (2003) *AGL24* acts as a promoter of flowering in *Arabidopsis* and is positively regulated by vernalization. *Plant J*, **33**, 867-874.
- Molinero-Rosales, N., Jamilena, M., Zurita, S., Gomez, P., Capel, J. and Lozano, R. (1999) *FALSIFLORA*, the tomato orthologue of *FLORICAULA* and *LEAFY*, controls flowering time and floral meristem identity. *Plant J*, **20**, 685-693.
- Moon, J., Suh, S.S., Lee, H., Choi, K.R., Hong, C.B., Paek, N.C., Kim, S.G. and Lee, I. (2003) The *SOC1* MADS-box gene integrates vernalization and gibberellin signals for flowering in *Arabidopsis*. *Plant J*, **35**, 613-623.
- Mouradov, A., Cremer, F. and Coupland, G. (2002) Control of flowering time: interacting pathways as a basis for diversity. *Plant Cell*, **14**, S111-130.
- Mouradov, A., Glassick, T., Hamdorf, B., Murphy, L., Fowler, B., Marla, S. and Teasdale, R.D. (1998) *NEEDLY*, a *Pinus radiata* ortholog of *FLORICAULA/LEAFY* genes, expressed in both reproductive and vegetative meristems. *PNAS*, **95**, 6537-6542.
- Mouradov, A., Hamdorf, B., Teasdale, R.D., Kim, J.T., Winter, K.U. and Theissen, G. (1999) A *DEF/GLO*-like MADS-box gene from a Gymnosperm: *Pinus radiata* contains an ortholog of Angiosperm B class floral homeotic genes. *Dev Genet*, **25**, 245-252.
- Mumberg, D., Muller, R. and Funk, M. (1995) Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds. *Gene*, **156**, 119-122.
- Nagasaki, H., Matsuoka, M. and Sato, Y. (2005) Members of TALE and WUS subfamilies of homeodomain proteins with potentially important functions in development form dimers within each subfamily in rice. *Genes Genet Syst*, **80**, 261-267.
- Nelson, P.C., Zurla, C., Brogioli, D., Beausang, J.F., Finzi, L. and Dunlap, D. (2006) Tethered particle motion as a diagnostic of DNA tether length. *J Phys Chem B*, **110**, 17260-17267.
- Ng, M. and Yanofsky, M.F. (2001) Function and evolution of the plant MADS-box gene family. *Nat Rev Genet*, **2**, 186-195.
- Ni, W., Xie, D., Hobbie, L., Feng, B., Zhao, D., Akkara, J. and Ma, H. (2004) Regulation of flower development in *Arabidopsis* by SCF complexes. *Plant Physiol*, **134**, 1574-1585.
- Niedziela-Majka, A., Rymarczyk, G., Kochman, M. and Ozyhar, A. (1998) GST-Induced dimerization of DNA-binding domains alters characteristics of their interaction with DNA. *Protein Expr Purif*, **14**, 208-220.
- Nielsen, A.D., Fuglsang, C.C. and Westh, P. (2003) Isothermal titration calorimetric procedure to determine protein-metal ion binding parameters in the presence of excess metal ion or chelator. *Anal Biochem*, **314**, 227-234.
- Nilsson, O., Lee, I., Blazquez, M.A. and Weigel, D. (1998) Flowering-time genes modulate the response to *LEAFY* activity. *Genetics*, **150**, 403-410.
- Onouchi, H., Igono, M.I., Perilleux, C., Graves, K. and Coupland, G. (2000) Mutagenesis of plants overexpressing *CONSTANS* demonstrates novel interactions among *Arabidopsis* flowering-time genes. *Plant Cell*, **12**, 885-900.

- Orchard, K. and May, G.E. (1993) An EMSA-based method for determining the molecular weight of a protein-DNA complex. *Nucleic Acids Res*, **21**, 3335-3336.
- Panne, D., Maniatis, T. and Harrison, S.C. (2004) Crystal structure of ATF-2/c-Jun and IRF-3 bound to the interferon-beta enhancer. *Embo J*, **23**, 4384-4393.
- Parcy, F. (2005) Flowering: a time for integration. *Int J Dev Biol*, **49**, 585-593.
- Parcy, F., Bomblies, K. and Weigel, D. (2002) Interaction of *LEAFY*, *AGAMOUS* and *TERMINAL FLOWER1* in maintaining floral meristem identity in *Arabidopsis*. *Development*, **129**, 2519-2527.
- Parcy, F., Nilsson, O., Bush, M.A., Lee, I. and Weigel, D. (1998) A genetic framework for floral patterning. *Nature*, **395**, 561-566.
- Parks, T.D., Howard, E.D., Wolpert, T.J., Arp, D.J. and Dougherty, W.G. (1995) Expression and purification of a recombinant tobacco etch virus Nla proteinase: biochemical analyses of the full-length and a naturally occurring truncated proteinase form. *Virology*, **210**, 194-201.
- Pelaz, S., Ditta, G.S., Baumann, E., Wisman, E. and Yanofsky, M.F. (2000) B and C floral organ identity functions require *SEPALLATA* MADS-box genes. *Nature*, **405**, 200-203.
- Pelaz, S., Gustafson-Brown, C., Kohalmi, S.E., Crosby, W.L. and Yanofsky, M.F. (2001) *APETALA1* and *SEPALLATA3* interact to promote flower development. *Plant J*, **26**, 385-394.
- Pineiro, M., Gomez-Mena, C., Schaffer, R., Martinez-Zapater, J.M. and Coupland, G. (2003) *EARLY BOLTING IN SHORT DAYS* is related to chromatin remodeling factors and regulates flowering in *Arabidopsis* by repressing *FT*. *Plant Cell*, **15**, 1552-1562.
- Putterill, J., Robson, F., Lee, K., Simon, R. and Coupland, G. (1995) The *CONSTANS* gene of *Arabidopsis* promotes flowering and encodes a protein showing similarities to zinc finger transcription factors. *Cell*, **80**, 847-857.
- Ratcliffe, O.J., Amaya, I., Vincent, C.A., Rothstein, S., Carpenter, R., Coen, E.S. and Bradley, D.J. (1998) A common mechanism controls the life cycle and architecture of plants. *Development*, **125**, 1609-1615.
- Ratcliffe, O.J., Bradley, D.J. and Coen, E.S. (1999) Separation of shoot and floral identity in *Arabidopsis*. *Development*, **126**, 1109-1120.
- Rezania, V., Tuszyński, J. and Hendzel, M. (2007) Modeling transcription factor binding events to DNA using a random walker/jumper representation on a 1D/2D lattice with different affinity sites. *Phys Biol*, **4**, 256-267.
- Riechmann, J.L., Heard, J., Martin, G., Reuber, L., Jiang, C., Keddie, J., Adam, L., Pineda, O., Ratcliffe, O.J., Samaha, R.R., Creelman, R., Pilgrim, M., Broun, P., Zhang, J.Z., Ghandehari, D., Sherman, B.K. and Yu, G. (2000) *Arabidopsis* transcription factors: genome-wide comparative analysis among eukaryotes. *Science*, **290**, 2105-2110.
- Riechmann, J.L., Krizek, B.A. and Meyerowitz, E.M. (1996) Dimerization specificity of *Arabidopsis* MADS domain homeotic proteins *APETALA1*, *APETALA3*, *PISTILLATA*, and *AGAMOUS*. *PNAS*, **93**, 4793-4798.
- Riese, M., Hohmann, S., Saedler, H., Munster, T. and Huiser, P. (2007) Comparative analysis of the *SBP*-box gene families in *P. patens* and seed plants. *Gene*, **401**, 28-37.
- Robles, P. and Pelaz, S. (2005) Flower and fruit development in *Arabidopsis thaliana*. *Int J Dev Biol*, **49**, 633-643.
- Roldan, M., Gomez-Mena, C., Ruiz-Garcia, L., Salinas, J. and Martinez-Zapater, J.M. (1999) Sucrose availability on the aerial part of the plant promotes morphogenesis and flowering of *Arabidopsis* in the dark. *Plant J*, **20**, 581-590.
- Rottmann, W.H., Meilan, R., Sheppard, L.A., Brunner, A.M., Skinner, J.S., Ma, C., Cheng, S., Jouanin, L., Pilate, G. and Strauss, S.H. (2000) Diverse effects of overexpression of *LEAFY* and *PTLF*, a poplar (*Populus*) homolog of *LEAFY/FLORICAULA*, in transgenic poplar and *Arabidopsis*. *Plant J*, **22**, 235-245.
- Ruiz-Garcia, L., Madueno, F., Wilkinson, M., Haughn, G., Salinas, J. and Martinez-Zapater, J.M. (1997) Different roles of flowering-time genes in the activation of floral initiation genes in *Arabidopsis*. *Plant Cell*, **9**, 1921-1934.
- Samach, A. and Coupland, G. (2000) Time measurement and the control of flowering in plants. *Bioessays*, **22**, 38-47.
- Samach, A., Klenz, J.E., Kohalmi, S.E., Risseuw, E., Haughn, G.W. and Crosby, W.L. (1999) The *UNUSUAL FLORAL ORGANS* gene of *Arabidopsis thaliana* is an F-box protein required for normal patterning and growth in the floral meristem. *Plant J*, **20**, 433-445.
- Samach, A., Onouchi, H., Gold, S.E., Ditta, G.S., Schwarz-Sommer, Z., Yanofsky, M.F. and Coupland, G. (2000) Distinct roles of *CONSTANS* target genes in reproductive development of *Arabidopsis*. *Science*, **288**, 1613-1616.
- Schmid, M., Uhlenhaut, N.H., Godard, F., Demar, M., Bressan, R., Weigel, D. and Lohmann, J.U. (2003) Dissection of floral induction pathways using global expression analysis. *Development*, **130**, 6001-6012.
- Schonrock, N., Bouveret, R., Leroy, O., Borghi, L., Kohler, C., Gruissem, W. and Hennig, L. (2006) Polycomb-group proteins repress the floral activator *AGL19* in the *FLC*-independent vernalization pathway. *Genes Dev*, **20**, 1667-1678.
- Schrag, J.D., Jiralerspong, S., Banville, M., Jaramillo, M.L. and O'Connor-McCourt, M.D. (2008) The crystal structure and dimerization interface of GADD45gamma. *PNAS*, **105**, 6566-6571.
- Schreiter, E.R., Sintchak, M.D., Guo, Y., Chivers, P.T., Sauer, R.T. and Drennan, C.L. (2003) Crystal structure of the nickel-responsive transcription factor NikR. *Nat Struct Biol*, **10**, 794-799.
- Schultz, E.A. and Haughn, G.W. (1991) *LEAFY*, a homeotic gene that regulates inflorescence development in *Arabidopsis*. *Plant Cell*, **3**, 771-781.
- Schumacher, M.A., Miller, M.C., Grkovic, S., Brown, M.H., Skurray, R.A. and Brennan, R.G. (2002) Structural basis for cooperative DNA binding by two dimers of the multidrug-binding protein QacR. *Embo J*, **21**, 1210-1218.
- Searle, I., He, Y., Turck, F., Vincent, C., Fornara, F., Krober, S., Amasino, R.A. and Coupland, G. (2006) The transcription factor *FLC* confers a flowering response to vernalization by repressing meristem competence and systemic signaling in *Arabidopsis*. *Genes Dev*, **20**, 898-912.
- Senechal, D.F., Ross, J.B. and Laue, T.M. (1998) Analysis of protein and DNA-mediated contributions to cooperative assembly of protein-DNA complexes. *Methods*, **16**, 3-20.
- Sessions, A., Yanofsky, M.F. and Weigel, D. (2000) Cell-cell signaling and movement by the floral transcription factors *LEAFY* and *APETALA1*. *Science*, **289**, 779-782.

- Shannon, S. and Meeks-Wagner, D.R. (1991) A mutation in the *Arabidopsis* *TFL1* gene affects inflorescence meristem development. *Plant Cell*, **3**, 877-892.
- Sheldon, C.C., Burn, J.E., Perez, P.P., Metzger, J., Edwards, J.A., Peacock, W.J. and Dennis, E.S. (1999) The *FLF* MADS box gene: a repressor of flowering in *Arabidopsis* regulated by vernalization and methylation. *Plant Cell*, **11**, 445-458.
- Shiu, S.H., Shih, M.C. and Li, W.H. (2005) Transcription factor families have much higher expansion rates in plants than in animals. *Plant Physiol*, **139**, 18-26.
- Simon, R., Carpenter, R., Doyle, S. and Coen, E. (1994) *Fimbriata* controls flower development by mediating between meristem and organ identity genes. *Cell*, **78**, 99-107.
- Simon, R. and Coupland, G. (1996) *Arabidopsis* genes that regulate flowering time in response to day-length. *Sem. Cell Dev. Biol.*, **7**, 419-425.
- Simpson, G.G. and Dean, C. (2002) *Arabidopsis*, the Rosetta stone of flowering time? *Science*, **296**, 285-289.
- Simpson, G.G., Gendall, A.R. and Dean, C. (1999) When to switch to flowering. *Annu Rev Cell Dev Biol*, **15**, 519-550.
- Sines, C.C., McFail-Isom, L., Howerton, S.B., Van Derveer, D. and D., W.L. (2000) Cation mediate B-DNA conformational heterogeneity. *J. Am. Chem.*, 11048-11056.
- Sliwinski, M.K., White, M.A., Maizel, A., Weigel, D. and Baum, D.A. (2006) Evolutionary divergence of *LFY* function in the mustards *Arabidopsis thaliana* and *Leavenworthia crassa*. *Plant Mol Biol*, **62**, 279-289.
- Smyth, D.R., Bowman, J.L. and Meyerowitz, E.M. (1990) Early flower development in *Arabidopsis*. *Plant Cell*, **2**, 755-767.
- Soltis, D.E., Chanderbali, A.S., Kim, S., Buzgo, M. and Soltis, P.S. (2007) The ABC model and its applicability to basal Angiosperms. *Ann Bot (Lond)*, **100**, 155-163.
- Souer, E., van der Krol, A., Kloos, D., Spelt, C., Blik, M., Mol, J. and Koes, R. (1998) Genetic control of branching pattern and floral identity during *Petunia* inflorescence development. *Development*, **125**, 733-742.
- Southerton, S.G., Strauss, S.H., Olive, M.R., Harcourt, R.L., Decroocq, V., Zhu, X., Llewellyn, D.J., Peacock, W.J. and Dennis, E.S. (1998) *Eucalyptus* has a functional equivalent of the *Arabidopsis* floral meristem identity gene *LEAFY*. *Plant Mol Biol*, **37**, 897-910.
- Suarez-Lopez, P., Wheatley, K., Robson, F., Onouchi, H., Valverde, F. and Coupland, G. (2001) *CONSTANS* mediates between the circadian clock and the control of flowering in *Arabidopsis*. *Nature*, **410**, 1116-1120.
- Sung, S. and Amasino, R.M. (2004) Vernalization and epigenetics: how plants remember winter. *Curr Opin Plant Biol*, **7**, 4-10.
- Tanahashi, T., Sumikawa, N., Kato, M. and Hasebe, M. (2005) Diversification of gene function: homologs of the floral regulator *FLO/LFY* control the first zygotic cell division in the moss *Physcomitrella patens*. *Development*, **132**, 1727-1736.
- Tandre, K., Albert, V.A., Sundas, A. and Engstrom, P. (1995) Conifer homologues to genes that control floral development in Angiosperms. *Plant Mol Biol*, **27**, 69-78.
- Tandre, K., Svenson, M., Svensson, M.E. and Engstrom, P. (1998) Conservation of gene structure and activity in the regulation of reproductive organ development of conifers and Angiosperms. *Plant J*, **15**, 615-623.
- Teper-Bamnolker, P. and Samach, A. (2005) The flowering integrator FT regulates *SEPALLATA3* and *FRUITFULL* accumulation in *Arabidopsis* leaves. *Plant Cell*, **17**, 2661-2675.
- Theissen, G. and Melzer, R. (2007) Molecular mechanisms underlying origin and diversification of the angiosperm flower. *Ann. Bot. (London)*, **100**, 603-619.
- Thingnaes, E., Torre, S., Ernstsén, A. and Moe, R. (2003) Day and night temperature responses in *Arabidopsis*: effects on gibberellin and auxin content, cell size, morphology and flowering time. *Ann Bot (Lond)*, **92**, 601-612.
- Tilly, J.J., Allen, D.W. and Jack, T. (1998) The *CArG* boxes in the promoter of the *Arabidopsis* floral organ identity gene *APETALA3* mediate diverse regulatory effects. *Development*, **125**, 1647-1657.
- Triezenberg, S.J., Kingsbury, R.C. and McKnight, S.L. (1988) Functional dissection of VP16, the *trans*-activator of *herpes simplex* virus immediate early gene expression. *Genes Dev.*, **2**, 718-729.
- Valverde, F., Mouradov, A., Soppe, W., Ravenscroft, D., Samach, A. and Coupland, G. (2004) Photoreceptor regulation of *CONSTANS* protein in photoperiodic flowering. *Science*, **303**, 1003-1006.
- van Pouderoyen, G., Ketting, R.F., Perrakis, A., Plasterk, R.H. and Sixma, T.K. (1997) Crystal structure of the specific DNA-binding domain of Tc3 transposase of *C.elegans* in complex with transposon DNA. *Embo J*, **16**, 6044-6054.
- Vazquez-Lobo, A., Carlsbecker, A., Vergara-Silva, F., Alvarez-Buylla, E.R., Pinero, D. and Engstrom, P. (2007) Characterization of the expression patterns of *LEAFY/FLORICAULA* and *NEEDLY* orthologs in female and male cones of the conifer genera *Picea*, *Podocarpus*, and *Taxus*: implications for current evo-devo hypotheses for Gymnosperms. *Evol Dev*, **9**, 446-459.
- Vinson, C., Myakishev, M., Acharya, A., Mir, A.A., Moll, J.R. and Bonovich, M. (2002) Classification of human B-ZIP proteins based on dimerization properties. *Mol Cell Biol*, **22**, 6321-6335.
- von Hippel, P.H. and Berg, O.G. (1989) Facilitated target location in biological systems. *J Biol Chem*, **264**, 675-678.
- Wachsstock, D.H. and Pollard, T.D. (1994) Transient state kinetics tutorial using the kinetics simulation program, KINSIM. *Biophys J*, **67**, 1260-1273.
- Wada, M., Cao, Q.F., Kotoda, N., Soejima, J. and Masuda, T. (2002) Apple has two orthologues of *FLORICAULA/LEAFY* involved in flowering. *Plant Mol Biol*, **49**, 567-577.
- Wagner, D., Sablowski, R.W. and Meyerowitz, E.M. (1999) Transcriptional activation of *APETALA1* by *LEAFY*. *Science*, **285**, 582-584.
- Wagner, D., Wellmer, F., Dilks, K., William, D., Smith, M.R., Kumar, P.P., Riechmann, J.L., Greenland, A.J. and Meyerowitz, E.M. (2004) Floral induction in tissue culture: a system for the analysis of *LEAFY*-dependent gene regulation. *Plant J*, **39**, 273-282.
- Wang, H., Chen, J., Wen, J., Tadege, M., Li, G., Liu, Y., Mysore, K.S., Ratet, P. and Chen, R. (2008) Control of compound leaf development by *FLORICAULA/LEAFY* ortholog *SINGLE LEAFLET1* in *Medicago truncatula*. *Plant Physiol*, **146**, 1759-1772.
- Wang, X., Feng, S., Nakayama, N., Crosby, W.L., Irish, V., Deng, X.W. and Wei, N. (2003) The COP9 signalosome interacts with SCF UFO and

- participates in *Arabidopsis* flower development. *Plant Cell*, **15**, 1071-1082.
- Weigel, D., Alvarez, J., Smyth, D.R., Yanofsky, M.F. and Meyerowitz, E.M. (1992) *LEAFY* controls floral meristem identity in *Arabidopsis*. *Cell*, **69**, 843-859.
- Weigel, D. and Meyerowitz, E.M. (1993a) Activation of floral homeotic genes in *Arabidopsis*. *Science*, **261**, 1723-1726.
- Weigel, D. and Meyerowitz, E.M. (1993b) *LEAFY* controls meristem identity in *Arabidopsis*. In Amasino, R. (ed.), *Cellular Communications in Plants*. Plenum Press, New York, pp. 115-122.
- Weigel, D. and Nilsson, O. (1995) A developmental switch sufficient for flower initiation in diverse plants. *Nature*, **377**, 495-500.
- Wenkel, S., Turck, F., Singer, K., Gissot, L., Le Gourrierec, J., Samach, A. and Coupland, G. (2006) CONSTANS and the CCAAT box binding complex share a functionally important domain and interact to regulate flowering of *Arabidopsis*. *Plant Cell*, **18**, 2971-2984.
- Whittaker, M.M., Mizuno, K., Bachinger, H.P. and Whittaker, J.W. (2006) Kinetic analysis of the metal binding mechanism of *Escherichia coli* manganese superoxide dismutase. *Biophys J*, **90**, 598-607.
- Wigge, P.A., Kim, M.C., Jaeger, K.E., Busch, W., Schmid, M., Lohmann, J.U. and Weigel, D. (2005) Integration of spatial and temporal information during floral induction in *Arabidopsis*. *Science*, **309**, 1056-1059.
- Wilkinson, M.D. and Haughn, G.W. (1995) *UNUSUAL FLORAL ORGANS* controls meristem identity and organ primordia fate in *Arabidopsis*. *Plant Cell*, **7**, 1485-1499.
- William, D.A., Su, Y., Smith, M.R., Lu, M., Baldwin, D.A. and Wagner, D. (2004) Genomic identification of direct target genes of *LEAFY*. *PNAS*, **101**, 1775-1780.
- Wilson, R.N., Heckman, J.W. and Somerville, C.R. (1992) Gibberellin is required for flowering in *Arabidopsis thaliana* under short days. *Plant Physiol.*, **100**, 403-408.
- Wu, X., Dinneny, J.R., Crawford, K.M., Rhee, Y., Citovsky, V., Zambryski, P.C. and Weigel, D. (2003) Modes of intercellular transcription factor movement in the *Arabidopsis* apex. *Development*, **130**, 3735-3745.
- Xu, H.E., Rould, M.A., Xu, W., Epstein, J.A., Maas, R.L. and Pabo, C.O. (1999) Crystal structure of the human Pax6 paired domain-DNA complex reveals specific roles for the linker region and carboxy-terminal subdomain in DNA binding. *Genes Dev*, **13**, 1263-1275.
- Xue, G.P. (2005) A CELF-fusion method for rapid determination of the DNA-binding sequence specificity of novel plant DNA-binding proteins. *Plant J*, **41**, 638-649.
- Yamasaki, K., Kigawa, T., Inoue, M., Tateno, M., Yamasaki, T., Yabuki, T., Aoki, M., Seki, E., Matsuda, T., Nunokawa, E., Ishizuka, Y., Terada, T., Shirouzu, M., Osanai, T., Tanaka, A., Seki, M., Shinozaki, K. and Yokoyama, S. (2004) A novel zinc-binding motif revealed by solution structures of DNA-binding domains of *Arabidopsis* SBP-family transcription factors. *J Mol Biol*, **337**, 49-63.
- Yamasaki, K., Kigawa, T., Inoue, M., Watanabe, S., Tateno, M., Seki, M., Shinozaki, K. and Yokoyama, S. (2008) Structures and evolutionary origins of plant-specific transcription factor DNA-binding domains. *Plant Physiol Biochem*, **46**, 394-401.
- Yasmin, R., Yeung, K.T., Chung, R.H., Gaczynska, M.E., Osmulski, P.A. and Noy, N. (2004) DNA-looping by RXR tetramers permits transcriptional regulation "at a distance". *J Mol Biol*, **343**, 327-338.
- Yokoyama, H., Fujii, S. and Matsui, I. (2008) Crystal structure of a core domain of stomatin from *Pyrococcus horikoshii* illustrates a novel trimeric and coiled-coil fold. *J Mol Biol*, **376**, 868-878.
- Yoo, S.K., Chung, K.S., Kim, J., Lee, J.H., Hong, S.M., Yoo, S.J., Yoo, S.Y., Lee, J.S. and Ahn, J.H. (2005) CONSTANS activates *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1* through *FLOWERING LOCUS T* to promote flowering in *Arabidopsis*. *Plant Physiol*, **139**, 770-778.
- York, D. and Reznikoff, W.S. (1996) Purification and biochemical analyses of a monomeric form of Tn5 transposase. *Nucleic Acids Res*, **24**, 3790-3796.
- Yu, H., Ito, T., Wellmer, F. and Meyerowitz, E.M. (2004) Repression of *AGAMOUS-LIKE 24* is a crucial step in promoting flower development. *Nat Genet*, **36**, 157-161.
- Yu, H., Xu, Y., Tan, E.L. and Kumar, P.P. (2002) *AGAMOUS-LIKE 24*, a dosage-dependent mediator of the flowering signals. *PNAS*, **99**, 16336-16341.
- Zahn, L.M., Leebens-Mack, J., DePamphilis, C.W., Ma, H. and Theissen, G. (2005) To B or Not to B a flower: the role of *DEFICIENS* and *GLOBOSA* orthologs in the evolution of the Angiosperms. *J Hered*, **96**, 225-240.
- Zahn, L.M., Leebens-Mack, J.H., Arrington, J.M., Hu, Y., Landherr, L.L., dePamphilis, C.W., Becker, A., Theissen, G. and Ma, H. (2006) Conservation and divergence in the *AGAMOUS* subfamily of MADS-box genes: evidence of independent sub- and neofunctionalization events. *Evol Dev*, **8**, 30-45.
- Zhang, P., Tan, H.T., Pwee, K.H. and Kumar, P.P. (2004) Conservation of class C function of floral organ development during 300 million years of evolution from Gymnosperms to Angiosperms. *Plant J*, **37**, 566-577.
- Zhao, D., Yang, M., Solava, J. and Ma, H. (1999) The *ASK1* gene regulates development and interacts with the *UFO* gene to control floral organ identity in *Arabidopsis*. *Dev Genet*, **25**, 209-223.
- Zimmer, J., Li, W. and Rapoport, T.A. (2006) A novel dimer interface and conformational changes revealed by an X-ray structure of *B. subtilis* SecA. *J Mol Biol*, **364**, 259-265.

Annexe 1

Article soumis à *EMBO J.*

Structural basis for LEAFY floral switch function and similarity with helix-turn-helix proteins

Cécile Hamès^{1,*}, Denis Ptchelkine^{2,3,*}, Clemens Grimm^{2,6}, Emmanuel Thevenon¹, Edwige Moyroud¹, Francine Gérard⁴, Jean-Louis Martiel⁵, Reyes Benlloch^{1,7}, François Parcy^{1,#} and Christoph W. Müller^{2,3,#}

¹Laboratoire Physiologie Cellulaire Végétale, UMR5168, Centre National de la Recherche Scientifique, Commissariat à l'énergie atomique, Institut National de la Recherche Agronomique, Université Joseph Fourier, 17 av. des Martyrs, bât. C2, 38054 Grenoble, France.

²European Molecular Biology Laboratory, Grenoble Outstation, B.P. 181, 38042 Grenoble, France.

³European Molecular Biology Laboratory, Structural and Computational Biology Unit, Meyerhofstrasse 1, 69117 Heidelberg, Germany.

⁴Unit of Virus Host Cell Interactions, UMR 5233 Université Joseph Fourier - European Molecular Biology Laboratory - Centre National de la Recherche Scientifique, Carl-Ivar Brändén Building, 6 rue Jules Horowitz, B.P. 181, 38042 Grenoble, France.

⁵ Université Joseph Fourier, TIMC-IMAG Laboratory, CNRS UMR 5525, INSERM, IRF 130 Grenoble, France.

⁶Present address: Institut für Biochemie, Biozentrum der Universität, Würzburg, Am Hubland, D-97074 Würzburg, Germany.

⁷Present address: Umeå Plant Science Centre, Department of Forest Genetics and Plant Physiology, Swedish University of Agricultural Sciences, SE-90183 Umeå, Sweden.

* These authors contributed equally to this work.

Corresponding authors: Christoph W. Müller (christoph.mueller@embl.de (C.W.M.)) and François Parcy (francois.parcy@cea.fr (F.P.))

Running Title: Structure of LEAFY/DNA complex

ABSTRACT

The LEAFY (LFY) protein is a key regulator of flower development in angiosperms. Its gradually increased expression governs the sharp floral transition and LFY subsequently controls the patterning of flower meristems by inducing the expression of floral homeotic genes. Despite a wealth of genetic data, how LFY functions at the molecular level is poorly understood. Here, we report crystal structures for the DNA-binding domain of *Arabidopsis thaliana* LFY bound to two target promoter elements. LFY adopts a novel seven-helix fold that binds DNA as a cooperative dimer, forming base-specific contacts in both the major and minor grooves. Cooperativity is mediated by two basic residues and plausibly accounts for LFY's effectiveness in triggering sharp developmental transitions. Our structure reveals an unexpected similarity between LFY and helix-turn-helix (HTH) proteins including homeodomain proteins known to regulate morphogenesis in higher eukaryotes. The appearance of flowering plants has been linked to the molecular evolution of LFY. Our work provides a unique framework to elucidate the molecular mechanisms underlying floral development and the evolutionary history of flowering plants.

Keywords: Crystal structure/ Flower development/ Homeotic genes / LEAFY/ Transcriptional regulation/

INTRODUCTION

Homeotic genes control developmental patterns and organ morphogenesis. In animals, they encode transcription factors of the homeodomain family, such as Hox and paired proteins, which contact DNA through one or several helix-turn-helix (HTH) motifs (Gehring et al., 1994; Underhill, 2000). In plants, most homeotic genes determining the identity of floral organs encode MADS-box transcription factors suggesting that plants and animals have adopted distinct types of homeotic regulators (Meyerowitz, 1997; Ng and Yanofsky, 2001). In addition to organ identity genes, plants also use another class of regulators named “meristem identity genes”, which control floral meristem versus shoot/inflorescence fate. In *Arabidopsis thaliana*, the meristem identity genes *LEAFY* (*LFY*) and *APETALA1* (*AP1*) induce flower development, whereas *TERMINAL FLOWER1* (*TFL1*) promotes inflorescence development (Blazquez et al., 2006). Mutations or ectopic expression of these genes result in complete or partial interconversions between flower and inflorescence meristems.

The *LFY* gene encodes a plant specific transcription factor, which plays a cardinal role in this process, regulating both the transition to flowering and the subsequent patterning of young floral meristems. During the plant vegetative growth, *LFY* expression increases in newly formed leaves until a certain threshold is reached: *LFY* then induces the expression of *AP1* and *CAULIFLOWER* (*CAL*) genes and triggers the abrupt floral transition (Blazquez et al., 2006). Once the floral meristem is established, *LFY* governs its spatial patterning by inducing the expression of the floral homeotic ABC genes, such as *AP1*, *APETALA3* (*AP3*) or *AGAMOUS* (*AG*), which control the identity of stereotypically arranged floral organs (Coen and Meyerowitz, 1991; Lohmann and Weigel, 2002).

LFY is found in all terrestrial plants from moss to angiosperms; its sequence shows a high level of conservation throughout the plant kingdom but no apparent similarity to other proteins (Maizel et al., 2005). Unlike many plant transcription factors that evolved by gene duplication to form a multigene family (Riechmann and Ratcliffe, 2000; Shiu et al., 2005), *LFY* is present in single copy in most angiosperms and *LFY* mutants available from several species such as snapdragon, petunia, tomato or maize show, as in *Arabidopsis*, partial or complete flower-to-shoot conversions (Bomblies et al., 2003; Coen et al., 1990; Molinero-Rosales et al., 1999; Souer et al., 1998). In gymnosperms, a paralogous *NEEDLY* (*NLY*) clade of genes exists. No mutant is available in these species, but *LFY* and *NLY* expression patterns are also consistent with a role in reproductive organ development (reviewed in Frohlich and Chase, 2007). Because of its central role in determining floral meristem identity, and considering that *NLY* disappeared concomitantly with the appearance of flowers, *LFY* has been put at the center of different evolutionary scenarios that rationalize the appearance of the successful angiosperm group (Albert et al., 2002; Frohlich, 2003; Frohlich and Chase, 2007; Theissen and Melzer, 2007).

LFY activates gene expression by recognizing pseudo-palindromic sequence elements (CCANTG^T/G) in the promoters of its target genes, including *AP1* (one site) and *AG* (4 sites, *AG-I* to *AG-IV*) (Busch et al., 1999; Hong et al., 2003; Lamb et al., 2002; Lohmann et al., 2001; Parcy et al., 1998). *LFY* has two domains, a partially conserved N-terminal domain that is thought to contribute to transcriptional activation and a highly conserved C-terminal domain responsible for DNA binding (*LFY-C*) (Coen et al., 1990; Maizel et al., 2005). *LFY* acts synergistically with coregulators such as the *WUSCHEL* (*WUS*) homeodomain protein (Lenhard et al., 2001; Lohmann et al., 2001) or the *UFO* F-Box protein (Chae et al., 2008; Lee et al., 1997; Parcy et al., 1998).

In this study, we show that *LFY* binds DNA cooperatively as a dimer, a property shown to be essential to trigger developmental switches. The crystal structure of *LFY-C* bound to DNA reveals the molecular basis for sequence specific recognition and cooperative binding as well as an unexpected similarity of

LFY with HTH proteins such as homeodomain transcription factors. Our findings enable to formulate new hypotheses on the appearance of angiosperms in evolution.

RESULTS AND DISCUSSION

LFY-C dimerizes upon DNA binding

We produced the recombinant LFY DNA-binding domain (LFY-C, residues 223-424) and showed by size exclusion chromatography (SEC) that it is monomeric in the absence of DNA (Figure 1AB). In electrophoretic mobility shift assays (EMSA), LFY-C recognized a DNA probe bearing an *AP1* site as two distinct species: a major protein-DNA complex and a minor one of higher mobility (Figure 1C). Multi-angle laser light scattering coupled to SEC demonstrated that the major complex contained two LFY-C molecules per DNA duplex (Figure 1B). The homodimeric nature of LFY in this complex was confirmed by mixing untagged and GFP-tagged LFY-C and observing a single new species attributable to the formation of a LFY-C/GFP-LFY-C/DNA complex (Figure 1C). Using probes mutated in one half-site of the palindrome, we confirmed that the minor, high-mobility species corresponds to a single LFY-C monomer bound to DNA (Supplementary Figure 2).

Structure of the LFY DNA-binding domain bound to its DNA recognition site

In order to understand how LFY specifically recognizes its DNA target sequences, we crystallized LFY-C in complex with DNA. We solved the structure of LFY-C bound to two different LFY binding sites, *AP1* and *AG-I* at 2.1 Å and 2.3 Å resolution, respectively (Figures 2, 3AB, Table 1). The overall structure shows a LFY-C dimer bound to a pseudo-palindromic DNA duplex, where the LFY-C monomers are related by a crystallographic dyad. The DNA duplexes used for co-crystallization deviate from strict two-fold symmetry at the 5' ends and at base pairs (bp) ± 9 , ± 7 and ± 0 in the *AP1* site (5' end, bp ± 7 , ± 6 , ± 4 and ± 0 in the *AG-I* site). Nevertheless the pseudo-dyads of the DNA duplexes coincide with the crystallographic dyad, probably as a result of the random bimodal orientation of the DNA duplex around the dyad (see Materials and Methods). The resulting molecular averaging does not impair our interpretation of the protein/DNA interface. In the final 2FoFc electron density map, but also in the initial solvent-flattened SIRAS electron density map (Supplementary Figure 1), the sugar-phosphate backbone of the DNA is well defined and shows no evidence of conformational averaging. The density for palindromic DNA bases related is clearly defined, whereas the density at non-palindromic positions is consistent with the superposition of two different base pairs. Furthermore, all residues close to the DNA are clearly defined and we do not observe any diffuse density, which suggests that each monomer only undergoes minor changes to adapt to the slightly different half sites. Despite the differences between the *AP1* and *AG-I* binding sites (Figure 2B), both complex structures are very similar and can be superimposed with an r.m.s. distance of 0.55 Å for 163 C α and 19 phosphate atoms.

LFY-C (with residues 237-399 ordered in the crystal structure) adopts a compact fold that interacts principally with a single DNA half-site (Figures 3AB, 4AB). The fold is defined by two short β -strands followed by seven helices connected by short loops (Figures 3AB). The absence of any extended hydrophobic patches at its surface suggests that LFY-C represents an autonomous DNA-binding domain without a large interface to its N-terminal domain. Helices $\alpha 2$ and $\alpha 3$ define a helix-turn-helix (HTH) motif (Aravind et al., 2005), with helix $\alpha 3$ occupying the major groove and mediating most of the DNA contacts. The DNA in the complex adopts a B-DNA-like conformation exhibiting an overall bend of about 20 degrees (Figure 3C), which can be localized to two kinks of about 10 degrees at base pairs $\pm 2/\pm 3$. Both ends of the DNA duplex are AT-rich and the minor grooves are narrower compared to classical B-DNA. Narrowing of the minor groove is slightly more pronounced in the *AG-I* duplex than in *AP1*.

DNA-recognition in the major and minor groove

Sequence-specific contacts between LFY and the DNA involve both the minor and major grooves. Base-specific contacts in the major groove are formed by Asn291 and Lys307 in helices $\alpha 2$ and $\alpha 3$,

which together specify the two invariant guanines at positions ± 2 and ± 3 (Figures 4AB). Mutating either of these residues into alanine resulted in considerably lower DNA binding affinity (Figure 4C), while previous studies showed loss of binding when the corresponding base pairs were mutated (Busch et al., 1999; Parcy et al., 1998). The *Arabidopsis lfy-20* mutation (N306D) adjacent to Lys307 also leads to a reduced DNA binding affinity (Supplementary Figure 3) and a weak *lfy* phenotype *in planta* (Maizel et al., 2005; Weigel et al., 1992), presumably because the negatively charged aspartate interacts unfavorably with the DNA backbone (Figures 4AB).

Base-specific recognition in the minor groove is mediated by Arg237, which is the first ordered N-terminal residue in the crystal structure. At the *AP1* site, its side chain points towards A:T base pairs ± 8 and contacts the exocyclic O2 of thymine 8 and also the O2 of cytosine 7 in one half-site, or the O2 of cytosine -9 in the other half-site (Figures 4AB, Supplementary Figure 4). In the *AG-I* site T:A base pair 8 is replaced by A:T, and in the *LFY/AG-I* complex the Arg237 side chain adopts a different conformation, which allows it to recognize the thymine of the opposite strand (Supplementary Figure 4). The importance of this interaction is underscored by the presence of A:T or T:A base pairs at position 8 in all 12 confirmed half sites (Busch et al., 1999; Hong et al., 2003; Lamb et al., 2002; Lohmann et al., 2001; Parcy et al., 1998). The consensus LFY binding site is therefore more accurately defined as $T/ANNNNCCANTGT/GNNNNT/A$ (with the center of the pseudo-palindrome underlined). The Arg237 side chain is inserted into an AT-rich narrow minor groove (Figure 4AB), similar to that observed in the Hox homeodomain-Exd-DNA complex, where the narrow minor groove was shown to enhance the electrostatic interaction between DNA backbone and arginine side chain (Joshi et al., 2007). The R237A mutation led to a strongly reduced affinity of LFY-C for *AP1* (Figure 4C). In contrast, changing the adenine 8 into a cytosine in *AP1* reduced only moderately the LFY-C binding affinity (Figure 4C, *AP1* m5), presumably because the arginine side chain can contact the adjacent base (Supplementary Figure 4). Finally, next to Arg237, the two *lfy* mutations (*lfy-4* (E238K), *lfy-5* (P240L)) result in decreased *in vitro* binding affinities (Supplementary Figure 3) and lead to a mutant phenotype *in planta* (Weigel et al., 1992).

An unusual contact with DNA is mediated by Pro308 that points between the guanines in base pairs ± 5 and ± 6 , which results in a pronounced propeller twist for base pair ± 5 and local bending of the DNA at this position (Figure 4B). The mutant *lfy-28* (P308L) is impaired in DNA binding and gives rise to an intermediate to strong phenotype *in planta* (Figures 4C-F), as a likely consequence of a steric clash of the leucine side chain with the guanine bases. In contrast, a small side chain like alanine perfectly fits in this protein/DNA interface and, indeed, the mutant protein P308A showed a wild-type binding affinity (Figure 4C). Pro308 is not strictly conserved and is substituted by serine in some *Brownea* species (Figure 2A). This substitution probably modifies DNA binding, because serine can form direct hydrogen bonds to DNA bases at positions ± 4 and ± 5 and it replaces P308, which locally distorts DNA. However, the conformational flexibility of serine and its ability to act as hydrogen bond donor or acceptor makes it difficult to predict the preferred binding specificity. Moreover, P308S is systematically associated with the K307R substitution, affecting the base-contacting residue Lys307 (Figure 2A). LFY proteins from *Brownea* species might therefore recognize significantly different DNA target sites.

Like in most protein/DNA co-crystal structures not all bases in the consensus LFY site $T/ANNNNCCANTGT/GNNNNT/A$ are specified through direct interactions with the protein. Additional specificity presumably arises from sequence-dependent deformability of the DNA, sometimes referred to as “indirect readout”. Dinucleotide steps CA/TG at bp ± 1 /bp ± 2 are part of the consensus LFY site and are particularly flexible, which might facilitate the observed kink of the DNA at base pairs $\pm 2/\pm 3$. However, these particular sequences are not critically required since they are not conserved in the *AP3-I* binding site (Lamb et al., 2002). “Indirect readout” also specifies the AT-rich sequences at both ends of the target site, where Arg237 inserts into a narrow minor groove.

Not all LFY mutations directly affect DNA contacts. Mutations *lfy-3* (T244M) and *lfy-9* (R331K) (Weigel et al., 1992) disturb two interacting amino acids, both of which contribute to a polar network which connects N-terminal residues with helices $\alpha 1$ and $\alpha 4$. Similarly, two other residues (His312 and Arg345) interact in a typical planar stacking. His312 and Arg345 are conserved except in the LFY protein from *Physcomitrella patens* (PpLFY1) where they are substituted by aspartate and cysteine, respectively (Maizel et al., 2005). His312 forms part of helix $\alpha 3$ and is located just one helical turn above Pro308 at the N-terminal end of helix $\alpha 3$. In addition, the preceding residue Lys307 directly contacts the guanine in base pairs ± 2 (Figure 4A). The loss of the His312/Arg345 stacking interaction in the moss PpLFY1 likely affects the orientation of helix $\alpha 3$, explaining the altered DNA binding properties of this ortholog, while reverting the aspartate into histidine restores the binding activity of PpLFY1 to canonical LFY binding sites (Maizel et al., 2005).

Structural basis for cooperative DNA binding

The structure of the LFY-C/DNA complex also reveals important monomer-monomer interactions governing its DNA binding mode (Figure 5A). Our EMSA analysis shows that LFY-C binds DNA in a cooperative manner: the monomeric complex is present only in minor amounts as compared to the dimeric complex, even at low LFY-C concentrations (Figure 5B) and binding of the second monomer occurs with a 90-fold higher affinity than binding of the first one (Figure 5C, Supplementary Figure 5). This type of cooperative binding can result either from DNA conformability, where binding of one monomer favors the binding of the second monomer, or from protein-protein interactions between DNA-bound monomers (Panne et al., 2004; Schumacher et al., 2002; Senear et al., 1998). Our structure suggests the latter. The LFY dimer comprises a small interface of 420 Å² buried surface area formed by loop $\alpha 12$ and helix $\alpha 7$ in which the two residues His387 and Arg390 form hydrogen bonds with the backbone carbonyl of Asp280 (Figure 5A). We validated the importance of these contacts by mutagenesis: cooperativity of binding is moderately affected in the H387A or R390A single mutants but more strongly reduced in a H387A/R390A double mutant (Figures 5BC, Supplementary Figure 5). Therefore, the small monomer-monomer interface (with a major contribution of His387 and Arg390) rather than DNA conformability is responsible for the cooperative binding. Whether the N-terminal domain of LFY also participates in dimerisation, in the presence or absence of DNA, will require additional experiments.

A better understanding of LFY's DNA binding mode also provides insight into its molecular switch function. DNA binding cooperativity, as well as dimerisation, allows transcription factors to work at lower concentrations and to enhance the sigmoidality of their response curves. When combined with feedback loops, it has been shown essential for threshold-dependent genetic switches (Burz et al., 1998; Cherry and Adler, 2000). LFY is involved in a positive autoregulation loop through activation of the homologous *AP1* and *CAL* genes, that in turn activate *LFY* expression (Bowman et al., 1993; Liljegren et al., 1999). LFY binding cooperativity combined with the AP1/CAL feedback loop therefore provides a plausible explanation for the threshold-dependent floral switch triggered by LFY.

Many transcription factors bind DNA as homodimers but also form heterodimers, thereby extending their spectrum of recognized DNA target sequences (Garvie and Wolberger, 2001; Klemm et al., 1998). LFY has been shown to activate the AG organ identity gene synergistically with the homeodomain protein WUS (Lohmann et al., 2001). Since adjacent WUS and LFY binding sites are present on the AG regulatory sequence, it has been suggested that LFY and WUS could bind simultaneously (Hong et al., 2003; Lohmann et al., 2001). Preliminary model building indicates that LFY homodimers cannot be accommodated with WUS at adjacent LFY and WUS binding sites. This observation raises the intriguing possibility that they might either compete for the same binding sites or more likely could form LFY-WUS heterodimers.

LFY shows similarities with HTH proteins

The nature and origin of LFY had so far remained elusive: LFY-C's primary sequence shows unusually strong sequence conservation within its family but has no apparent similarity to any described transcription factor. The crystal structure of LFY-C bound to DNA reveals a seven-helix domain with many residues involved in protein-DNA interactions, tightly constrained packing interactions in the hydrophobic core and protein/protein interactions with the other monomer. Presumably, these observed tight structural and functional constraints on many residues spread over the entire DNA-binding domain explain the high level of sequence conservation within LFY-C.

The LFY-C structure contains an unpredicted HTH motif formed by helices $\alpha 2$ and $\alpha 3$ as part of the overall fold. HTH motifs are present in a wide variety of DNA binding proteins throughout the three kingdoms of life. They are typically found in a bundle of three to six α -helices or combined with β sheets (winged HTH/fork head domain), which provide a stabilizing hydrophobic core (Aravind et al., 2005; Weigel and Jackle, 1990). Comparison of LFY-C against the Protein Data Bank using program DALI (Holm and Sander, 1993) detects similarity of relative short β -helical segments (~60 amino acid residues) with different α -helical proteins including HTH proteins (maximal Dali Z-score 3.0, pairs with Z < 2.0 are structurally dissimilar). A search comprising only the first three N-terminal helices including the HTH motif mainly showed similarity to different HTH proteins with slightly higher scores (maximal Dali Z-score: 4.5). When considering just the three helices $\alpha 1$, $\alpha 2$ and $\alpha 3$, LFY aligns well with other 3-helix bundle HTH proteins including the homeodomain protein engrailed (r.m.s.d._{40C α} = 2.9 Å), the paired domain (r.m.s.d._{44C α} = 3.5 Å) and the Tc3 transposase (r.m.s.d._{30C α} = 2.4 Å). LFY and partitioning protein KorB (r.m.s.d._{71C α} = 3.7 Å) share some similarity beyond the typical DNA/RNA-binding 3-helical bundle core (Khare et al., 2004; Russell and Barton, 1992), where five of the seven LFY-C helices, including the HTH motif, roughly superimpose with KorB helices. However, LFY cannot easily be assigned to any of the described classes of HTH proteins (Aravind et al., 2005) and it therefore represents a new variant of multi-helical bundle proteins.

The DNA recognition mode of LFY is similar to those observed for the paired domain, Tc3A transposase, Hin recombinase, and λ repressor (van Pouderoyen et al., 1997; Xu et al., 1999). The axis of the recognition helix $\alpha 3$ in the HTH of these proteins is oriented parallel to the edges of the nucleotide bases. Only the N-terminus of the recognition helix is inserted in the major groove of the DNA, whereas the short helix $\alpha 2$ plays a supporting role. In contrast, in homeodomain proteins the long probe helix $\alpha 3$ runs more parallel to the neighboring DNA phosphate backbone, and mainly the central part of helix $\alpha 3$ contacts the DNA (Figure 6). Similarity between LFY and the paired domain also includes a small two-stranded β sheet which precedes the three-helix bundle and N-terminal residues which are inserted into the minor groove. However, the minor groove contacting residues are located at the most N-terminal end of LFY-C, while in the paired domain they protrude from the loop connecting the two short N-terminal β strands.

Sequence similarities are too weak to suggest a precise evolutionary origin for LFY, although structural resemblances indicate that it might derive from ancestral HTH proteins including paired and homeodomain proteins (Aravind et al., 2005; Breitling and Gerber, 2000; Rosinski and Atchley, 1999). Until now, most plant homeotic genes were found to encode MADS box transcription factors, whereas plant homeodomain proteins rather control meristem homeostasis and cell division (Meyerowitz, 1997; Ng and Yanofsky, 2001). Our study reveals that the LFY master regulator, which determines flower meristem fate and controls the expression of floral organ identity genes, shares structural similarity with other HTH proteins, indicating that this universal DNA binding motif has also been adopted in plants to trigger major developmental switches.

Prospects regarding the appearance of angiosperms.

The LFY-C structure combined with more than 200 LFY sequences from all types of terrestrial plants offers a unique opportunity to detect key residues in evolution. Some charged LFY-C surface residues (such as Lys253 or Lys254) are strictly conserved, suggesting that they might participate in interactions with other proteins. Other residues are conserved in all angiosperms but not in the non-flowering plants. For example, R390, identified as one of the residues mediating interaction between monomers and cooperative binding, has been conserved in angiosperm LFY proteins, whereas most LFY from non-flowering plants, such as gymnosperms and ferns, show a lysine at this position. This amino acid change presumably weakens the interaction between monomers and thereby reduces the DNA binding affinity. The acquisition of R390 might therefore have been important for flower evolution. Because LFY stands at the very center of the network regulating flower development, it has been proposed that modifications of the *LFY* gene contributed to the appearance of floral structures in evolution (Albert et al., 2002; Frohlich, 2003; Frohlich and Chase, 2007; Theissen and Melzer, 2007). The availability of the LFY-C crystal structure provides a unique framework for generating plausible hypotheses that relate the appearance of angiosperms to specific events during the molecular evolution of LFY. The “functional synthesis” approach that combines phylogeny, biochemical and structural analyses with functional assays *in vivo* (Dean and Thornton, 2007) can now be applied to LFY to try to solve one of the most puzzling enigmas of plant biology: the origin of flowers.

MATERIALS AND METHODS

Plant material

The *lfy-28* mutant allele of *Arabidopsis thaliana* (accession Landsberg *erecta*) was kindly provided by D. Weigel (Max Planck Institute, Tübingen, Germany) and originally isolated by J. Fletcher (PGEN, Albany). *lfy-28* mutant had been back-crossed twice to the wild type, and individuals showing a mutant phenotype were selected from segregating populations. Plants were grown at 25°C in long days (16 hours light).

Plasmid constructions:

Expression plasmids: LFY-C (residues 223-424 from *Arabidopsis thaliana* *LEAFY* cDNA) was amplified from pIL-8 (obtained from D. Weigel) with Pfu Turbo Polymerase (Stratagene, France) and primers oFP1242 (5'CTCTCGAGCCCGGGCTAGAAACGCAAGTCGTCGCC3') and oFP1244 (5'CTCTCGAGCCCGGGCTATCCGGTACAGCTAATACCGCC3'), subcloned into pCR-TOPO-BluntII (Invitrogen, Cergy Pontoise, France) and shuttled to pETM-11 (Dummler et al., 2005) as NcoI/XhoI fragment to yield the pCH28 expression vector. pETM-11 contains a N-terminal 6xHis-tag followed by a TEV cleavage site.

LFY-GFP plasmid: a GFP fragment was amplified from pBS-GLFY plasmid obtained from X. Wu (Wu et al., 2003) using primers oETH1001 5'CCCACTACTGAGAATCTTTATTTTCAGGGCCAGTTCAGTAAAGGAGAAGAAC3' and oETH1002 5'CCCCAAACCACTACCTCCGTTGCCGTTATCCTGTTGTATAGTTCATCCAT. The amplified fragment was subsequently used as a megaprimer to amplify plasmid pCH28 and yield pETH8 (6His-TEV-GFP-LFY-C).

Expression plasmids for mutant LFY-C: pCH45 (K307A), pCH46 (N291A), pCH47 (R237A), pCH48 (P308A), pCH49 (D280K), pCH50 (H387A R390A), pCH54 (H387A), pEDW127 (R390A), pCH55 (*lfy-28*, P308L), pETH21 (*lfy-4*, E238K), pETH23 (*lfy-20*, N306D), and pCH56 (*lfy-5*, P240L) were derived from pCH28 using the megaprimer strategy with appropriate primers (Kirsch and Joly, 1998). All plasmids were verified by sequencing.

Protein Expression, Purification and Crystallization

Wild type and mutant LFY-C domains were expressed using *Escherichia coli* strain RosettaBlue(DE3)pLysS (Novagen, Strasbourg, France). After induction by 0.5 mM IPTG cells were grown overnight at 22°C. For cell lysis, the pellet of 1 liter culture was sonicated in 30 mL lysis buffer A (500 mM NaCl, 20 mM Tris-HCl pH 8, 5 mM imidazole, 5 % glycerol, 5 mM Tris(2-carboxyethyl)phosphine hydrochloride (TCEP)), 1 protease inhibitor cocktail tablet Complete EDTA-free (Roche, Meylan, France) and centrifuged for 40 min at 30000 g. The clear supernatant was incubated for about 1 hour with 1 ml Ni-NTA resin (Qiagen, Courtaboeuf, France). The resin was transferred into a column, washed with 20 column volumes (CV) of buffer A, buffer A + 50 mM imidazole (10 CV) and eluted with buffer A + 380 mM imidazole. The fractions containing the protein were pooled and applied to a Hi-load Superdex-200 16/60 prep grade column (GE Healthcare, Orsay, France) equilibrated with 200 mM NaCl, 20 mM Tris-HCl pH 8, 5 mM dithiothreitol (DTT) to eliminate aggregated proteins by size exclusion chromatography (SEC). Protein concentration was estimated using the Bradford assay (Bradford, 1976).

For crystallographic experiments, after elution on the metal affinity column, the histidine tag was cleaved at 4 °C overnight with tobacco etch virus protease (TEV; 0.01 % w/w, 16 hours, 4 °C) during the dialysis step against buffer B (500 mM NaCl, 20 mM Tris pH 7.5, 5 mM DTT). The TEV protease, the histidine tag and the uncleaved protein were removed by repassing

the dialyzed sample over the Ni-affinity column. The protein was separated from the remaining DNA contamination using the anion-exchange column MonoQ HR10/10 (GE Healthcare, Orsay, France) pre-equilibrated in buffer B. Pure protein was recovered in the flow-through while DNA remained bound to the resin. Aggregated protein was removed by size-exclusion chromatography with Superdex S75GL column (GE Healthcare, Orsay, France) in 200 mM NaCl, 10 mM Tris pH 7.5, 5 mM DTT. The protein concentration was adjusted to 7.5 mg/ml. DNA oligonucleotides were chemically synthesized and purified by anion-exchange chromatography following established procedures (Cramer and Muller, 1997).

Electrophoretic mobility shift assays (EMSA)

Single stranded oligonucleotides, 5' labeled with tetra-methylcarboxy-rhodamine (TAMRA, Sigma, Saint Quentin Fallavier, France) were annealed to non-fluorescent complementary oligonucleotides in annealing buffer (10 mM Tris pH 7.5, 150 mM NaCl, 1 mM EDTA). The sequences of oligonucleotides used are indicated in Supplementary Table 1. Binding reactions were performed in 20 μ L binding buffer (150 mM NaCl, 20 mM Tris-HCl pH 7.5, 1 % glycerol, 0.25 mM EDTA, 2 mM $MgCl_2$, 1 mM DTT) supplemented with 28 ng/ μ L fish sperm DNA (Roche, Meylan, France) and 10 nM double stranded DNA probe or 140 ng/ μ L fish sperm DNA for 50 nM DNA probe (Figure 5). Binding reactions were loaded onto native 6 % polyacrylamide gels 0.5 X TBE (45 mM Tris, 45 mM boric acid and 1 mM EDTA pH 8) and electrophoresed at 90 V for 80 min at 4 °C. Gels were scanned on a Typhoon 9400 scanner (Molecular dynamics, Sunnyvale, CA. Excitation light 532 nm, Emission filter 580 BP 30) and signals were quantified using ImageQuant software (Molecular dynamics, Sunnyvale, CA). Estimations of K_d1 and K_d2 (Figure 5 and Supplementary Figure 5) were based on the quantifications of binding experiments shown in Figure 5B. The binding model equations used to calculate these K_d s are explained in detail in Supplementary Methods.

Size exclusion chromatography

The molecular size of LFY-C/AP1 complex was determined using a Superdex-200 10/300GL column (GE Healthcare, Orsay, France), equilibrated with buffer containing 150 mM NaCl, 16 mM Tris-HCl pH 7.5, 0.6 mM EDTA and 1 mM DTT, and calibrated with low and high molecular weight protein standards (Gel filtration calibration kit, GE Healthcare, Orsay, France). Our samples (LFY-C 40 μ M, AP1 WT 10 μ M, LFY-C 40 μ M + AP1 WT 10 μ M) were analyzed in the same buffer as protein standards, and molecular size is deduced from the standard curve.

Analytical Size Exclusion Chromatography and Multi-angle Laser Light Scattering (MALLS-SEC).

Separation by size exclusion chromatography (SEC) was carried out with a S200 Superdex column (GE Healthcare, Orsay, France). The column was equilibrated in 20 mM Tris-HCl, 150 mM NaCl buffer at pH 7.5. Separations were performed at 20 °C with a flow rate of 0.6 mL min⁻¹. 50 μ L of protein solution at a concentration of 5 mg mL⁻¹ were injected. The elution was monitored by using a DAWN-EOS detector with a laser emitting at 690 nm for on-line Multi-Angle Laser Light Scattering measurement (Wyatt Technology Corp., Santa Barbara, CA.), and with a RI2000 detector for on-line refractive index measurements (Schambeck SFD). Molecular mass calculation was performed as described using the ASTRA software (Gerard et al., 2007).

Crystallization

For co-crystallization using the hanging drop method, protein and DNA duplexes were mixed in a molar ratio 2:1. The best crystals were obtained at 4 °C with 20-mer oligonucleotides bearing complementary A:T overhangs and with 10% PEG 400, 100 mM KCl, 10 mM $CaCl_2$, 50 mM HEPES (NaOH) pH 7.0 as reservoir solution. Single crystals grew to a maximal size of 300×300×500 μ m³ and were stepwise transferred to reservoir solution containing 30 % (v/v) glycerol for cryo-protection. For preparation of the mercury derivative the crystals were soaked in the reservoir solution supplemented with 0.1 mM ethylmercury-thiosalicylate (EMTS) for two hours.

X-ray structure determination

The crystals of the LFY-C/AP1-DNA complex belong to space group P6₅22 ($a = b = 98.8$ Å, $c = 177.4$ Å), diffracted up to 2.1 Å resolution and contain half a complex per asymmetric unit. Crystals of the LFY-C/AG-I-DNA complex are isomorphous but diffracted slightly weaker (Table 1). Diffraction data collected at ESRF beamlines ID14-1, ID29 and ID23-2 were processed using program XDS (Kabsch, 1993). The structure of the LFY-C/AP1-DNA complex was solved using the single isomorphous replacement with anomalous scattering (SIRAS) method with EMTS as derivative. The quality of native and derivative data sets is summarized in Table 1. Mercury sites were located using program SOLVE (Terwilliger and Berendzen, 1999) and phases were calculated with program SHARP (de la Fortelle and Bricogne, 1997). The experimental electron density map (Supplementary Figure 1) allowed us to automatically build the initial model using program ARP/wARP (Perrakis et al., 2001) followed by manually adjusting some side chain conformations with program COOT (Emsley and Cowtan, 2004) and refinement with program Refmac5 including a TLS refinement with 7 groups (Murshudov et al., 1997) and later with program Phenix (Adams et al., 2002). In space group P6₅22, the two monomers bound to the pseudo-palindromic DNA duplex are related by a crystallographic dyad. In the crystal, the pseudo-dyad of the DNA coincides with the crystallographic dyad, although the DNA duplexes deviate from strict twofold symmetry at base pairs 0, ± 7 , ± 9 and the overhanging 5'-end, in the AP1 site and at base pairs 0, ± 4 , ± 6 and ± 7 and the overhanging 5'-end in the AG-I site. To confirm our space group assignment and the underlying assumption that the DNA duplexes used for co-crystallization are randomly distributed in two

orientations, the data were reprocessed in the lower symmetry space group $P6_5$ lacking the dyad, which did not significantly change the R_{meas} values. Subsequently, models of the LFY-C dimer bound to the 20-mer DNA duplex were built in space group $P6_5$ for the *AP1* and *AG-I* sites and refined in two independent orientations yielding very similar final R_{cryst} and R_{free} values compared to the refinement in space group $P6_522$. In both orientations (and for both target sites), the final Fo-Fc electron density maps showed pairs of difference fourier peaks ($\sim 7\sigma$) of similar height at the non-palindromic bases indicating that a unique orientation of the DNA duplexes does not correctly describe the situation in the crystals. Finally, simulated-annealing omit maps in space group $P6_5$ where the non-palindromic bases were omitted showed averaged densities for the omitted bases in both complexes, further confirming the assigned space group $P6_522$.

To account for the two orientations of the DNA in the crystal during the refinement, two nucleotides with 50% occupancy were introduced at the non-palindromic positions. The final model of the LFY-C/*AP1* complex at 2.1 Å resolution ($R_{\text{cryst}} = 21.0\%$; $R_{\text{free}} = 23.7\%$) comprises residues 237-399 of the LFY DNA binding domain, while the poorly conserved 25 C-terminal residues are disordered. For the refinement of the LFY/*AG-I* complex the *AP1* DNA sequence in the LFY/*AP1* complex was replaced with the *AG-I* sequence. Multiple rounds of TLS refinement (including TLS refinement with 7 groups) using program Refmac5 (Murshudov et al., 1997) and Phenix (Adams et al., 2002) yielded a model with R_{cryst} of 22.1 % and R_{free} of 24.9 % using data between 20 - 2.3 Å resolution. The atomic coordinates and structure factors for the LFY/*AP1* and LFY/*AG-I* complexes have been deposited with the Protein Data Bank under accession codes xyz1, xyz2, respectively.

ACKNOWLEDGEMENTS

We thank D. Weigel (MPI, Tübingen) for providing materials and advice, X. Wu for material, L. Blanchoin, C. Guérin, R. Dumas, M. Jamin, C. Ebel and G. Schoen for help with protein expression and characterization, L. Blanchoin, A. Maizel, M. Blazquez, E. Dorcey, and C. Petosa for critical reading of the manuscript, R. Russell for structure comparisons, J.-L. Martiel for binding modeling, the EMBL/ESRF Joint Structural Biology Group for access and support at the ESRF beamlines and the EMBL Grenoble crystallization facility for support. Funding by ATIP (CNRS) to FP and RB, ATIP+ (CNRS), ANR Plant-TF code to FP, Region Rhône-Alpes/cluster 9 to CH, Programme Emergence of the Region Rhône-Alpes for DP.

REFERENCES

- Adams, P.D., Grosse-Kunstleve, R.W., Hung, L.W., Ioerger, T.R., McCoy, A.J., Moriarty, N.W., Read, R.J., Sacchettini, J.C., Sauter, N.K. and Terwilliger, T.C. (2002) PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr*, **58**, 1948-1954.
- Albert, V.A., Oppenheimer, D.G. and Lindqvist, C. (2002) Pleiotropy, redundancy and the evolution of flowers. *Trends Plant Sci.*, **7**, 297-301.
- Aravind, L., Anantharaman, V., Balaji, S., Babu, M.M. and Iyer, L.M. (2005) The many faces of the helix-turn-helix domain: transcription regulation and beyond. *FEMS Microbiol Rev*, **29**, 231-262.
- Blazquez, M.A., Ferrandiz, C., Madueno, F. and Parcy, F. (2006) How floral meristems are built. *Plant Mol Biol*, **60**, 855-870.
- Bomblies, K., Wang, R.L., Ambrose, B.A., Schmidt, R.J., Meeley, R.B. and Doebley, J. (2003) Duplicate *FLORICAULA/LEAFY* homologs *zfl1* and *zfl2* control inflorescence architecture and flower patterning in maize. *Development*, **130**, 2385-2395.
- Bowman, J.L., Alvarez, J., Weigel, D., Meyerowitz, E.M. and Smyth, D.R. (1993) Control of Flower development in *Arabidopsis thaliana* by *APETALA1* and interacting genes. *Development*, **119**, 721-743.
- Bradford, M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem*, **72**, 248-254.
- Breitling, R. and Gerber, J.K. (2000) Origin of the paired domain. *Dev Genes Evol*, **210**, 644-650.
- Burz, D.S., Rivera-Pomar, R., Jackle, H. and Hanes, S.D. (1998) Cooperative DNA-binding by Bicoid provides a mechanism for threshold-dependent gene activation in the *Drosophila* embryo. *Embo J*, **17**, 5998-6009.
- Busch, M.A., Bomblies, K. and Weigel, D. (1999) Activation of a floral homeotic gene in *Arabidopsis*. *Science*, **285**, 585-587.
- Chae, E., Tan, Q.K., Hill, T.A. and Irish, V.F. (2008) An *Arabidopsis* F-box protein acts as a transcriptional co-factor to regulate floral development. *Development*, **135**, 1235-1245.
- Cherry, J.L. and Adler, F.R. (2000) How to make a biological switch. *J Theor Biol*, **203**, 117-133.
- Coen, E.S. and Meyerowitz, E.M. (1991) The war of the whorls: genetic interactions controlling flower development. *Nature*, **353**, 31-37.
- Murphy, G. and Carpenter, R. (1990) *floricaula*: a homeotic gene required for flower development in *Antirrhinum majus*. *Cell*, **63**, 1311-1322.
- Cramer, P. and Muller, C.W. (1997) Engineering of diffraction-quality crystals of the NF-kappaB P52 homodimer:DNA complex. *FEBS Lett*, **405**, 373-377.
- de la Fortelle, E. and Bricogne, G. (1997) Maximum-likelihood heavy-atom parameter refinement for the multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Meth Enzym*, **276**, 472-494.
- Dean, A.M. and Thornton, J.W. (2007) Mechanistic approaches to the study of evolution: the functional synthesis. *Nat Rev Genet*, **8**, 675-688.
- Delano, W.L. (2002) The PyMOL molecular graphics system. DeLano Scientific, Palo Alto, CA.
- Dummler, A., Lawrence, A.M. and de Marco, A. (2005) Simplified screening for the detection of soluble fusion constructs expressed in *E. coli* using a modular set of vectors. *Microb Cell Fact*, **4**, 34.
- Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr*, **60**, 2126-2132.
- Frohlich, M.W. (2003) An evolutionary scenario for the origin of flowers. *Nat Rev Genet*, **4**, 559-566.
- Frohlich, M.W. and Chase, M.W. (2007) After a dozen years of progress the origin of angiosperms is still a great mystery. *Nature*, **450**, 1184-1189.
- Garvie, C.W. and Wolberger, C. (2001) Recognition of specific DNA sequences. *Mol Cell*, **8**, 937-946.

- Gehring, W.J., Affolter, M. and Burglin, T. (1994) Homeodomain proteins. *Annu Rev Biochem*, **63**, 487-526.
- Gerard, F.C., Ribeiro Ede, A., Jr., Albertini, A.A., Gutsche, I., Zaccari, G., Ruigrok, R.W. and Jamin, M. (2007) Unphosphorylated rhabdoviridae phosphoproteins form elongated dimers in solution. *Biochemistry*, **46**, 10328-10338.
- Holm, L. and Sander, C. (1993) Protein structure comparison by alignment of distance matrices. *J Mol Biol*, **233**, 123-138.
- Hong, R.L., Hamaguchi, L., Busch, M.A. and Weigel, D. (2003) Regulatory elements of the floral homeotic gene *AGAMOUS* identified by phylogenetic footprinting and shadowing. *Plant Cell*, **15**, 1296-1309.
- Joshi, R., Passner, J.M., Rohs, R., Jain, R., Sosinsky, A., Crickmore, M.A., Jacob, V., Aggarwal, A.K., Honig, B. and Mann, R.S. (2007) Functional specificity of a Hox protein mediated by the recognition of minor groove structure. *Cell*, **131**, 530-543.
- Kabsch, W.J. (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J Appl Cryst*, **26**, 795-800.
- Khare, D., Ziegler, G., Lanka, E. and Heinemann, U. (2004) Sequence-specific DNA binding determined by contacts outside the helix-turn-helix motif of the ParB homolog KorB. *Nat Struct Mol Biol*, **11**, 656-663.
- Kirsch, R.D. and Joly, E. (1998) An improved PCR-mutagenesis strategy for two-site mutagenesis or sequence swapping between related genes. *Nucleic Acids Res*, **26**, 1848-1850.
- Klemm, J.D., Schreiber, S.L. and Crabtree, G.R. (1998) Dimerization as a regulatory mechanism in signal transduction. *Annu Rev Immunol*, **16**, 569-592.
- Lamb, R.S., Hill, T.A., Tan, Q.K. and Irish, V.F. (2002) Regulation of *APETALA3* floral homeotic gene expression by meristem identity genes. *Development*, **129**, 2079-2086.
- Lee, I., Wolfe, D.S., Nilsson, O. and Weigel, D. (1997) A *LEAFY* co-regulator encoded by *UNUSUAL FLORAL ORGANS*. *Curr Biol*, **7**, 95-104.
- Lenhard, M., Bohnert, A., Jurgens, G. and Laux, T. (2001) Termination of stem cell maintenance in Arabidopsis floral meristems by interactions between *WUSCHEL* and *AGAMOUS*. *Cell*, **105**, 805-814.
- Liljegren, S.J., Gustafson-Brown, C., Pinyopich, A., Ditta, G.S. and Yanofsky, M.F. (1999) Interactions among *APETALA1*, *LEAFY*, and *TERMINAL FLOWER1* specify meristem fate. *Plant Cell*, **11**, 1007-1018.
- Lohmann, J.U., Hong, R.L., Hobe, M., Busch, M.A., Parcy, F., Simon, R. and Weigel, D. (2001) A molecular link between stem cell regulation and floral patterning in Arabidopsis. *Cell*, **105**, 793-803.
- Lohmann, J.U. and Weigel, D. (2002) Building beauty: the genetic control of floral patterning. *Dev Cell*, **2**, 135-142.
- Maizel, A., Busch, M.A., Tanahashi, T., Perkovic, J., Kato, M., Hasebe, M. and Weigel, D. (2005) The floral regulator *LEAFY* evolves by substitutions in the DNA binding domain. *Science*, **308**, 260-263.
- Meyerowitz, E.M. (1997) Plants and the logic of development. *Genetics*, **145**, 5-9.
- Molinero-Rosales, N., Jamilena, M., Zurita, S., Gomez, P., Capel, J. and Lozano, R. (1999) *FALSIFLORA*, the tomato orthologue of *FLORICAULA* and *LEAFY*, controls flowering time and floral meristem identity. *Plant J*, **20**, 685-693.
- Murshudov, G.N., Vagin, A.A. and Dodson, E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr*, **53**, 240-255.
- Ng, M. and Yanofsky, M.F. (2001) Function and evolution of the plant MADS-box gene family. *Nat Rev Genet*, **2**, 186-195.
- Panne, D., Maniatis, T. and Harrison, S.C. (2004) Crystal structure of ATF-2/c-Jun and IRF-3 bound to the interferon-beta enhancer. *Embo J*, **23**, 4384-4393.
- Parcy, F., Nilsson, O., Bush, M.A., Lee, I. and Weigel, D. (1998) A genetic framework for floral patterning. *Nature*, **395**, 561-566.
- Perrakis, A., Harkiolaki, M., Wilson, K.S. and Lamzin, V.S. (2001) ARP/wARP and molecular replacement. *Acta Crystallogr D Biol Crystallogr*, **57**, 1445-1450.
- Riechmann, J.L. and Ratcliffe, O.J. (2000) A genomic perspective on plant transcription factors. *Curr Opin Plant Biol*, **3**, 423-434.
- Rosinski, J.A. and Atchley, W.R. (1999) Molecular evolution of helix-turn-helix proteins. *J Mol Evol*, **49**, 301-309.
- Russell, R.B. and Barton, G.J. (1992) Multiple protein sequence alignment from tertiary structure comparison: assignment of global and residue confidence levels. *Proteins*, **14**, 309-323.
- Schumacher, M.A., Miller, M.C., Grkovic, S., Brown, M.H., Skurray, R.A. and Brennan, R.G. (2002) Structural basis for cooperative DNA binding by two dimers of the multidrug-binding protein QacR. *Embo J*, **21**, 1210-1218.
- Senear, D.F., Ross, J.B. and Laue, T.M. (1998) Analysis of protein and DNA-mediated contributions to cooperative assembly of protein-DNA complexes. *Methods*, **16**, 3-20.
- Shiu, S.H., Shih, M.C. and Li, W.H. (2005) Transcription factor families have much higher expansion rates in plants than in animals. *Plant Physiol*, **139**, 18-26.
- Souer, E., van der Krol, A., Kloos, D., Spelt, C., Bliet, M., Mol, J. and Koes, R. (1998) Genetic control of branching pattern and floral identity during *Petunia* inflorescence development. *Development*, **125**, 733-742.
- Terwilliger, T.C. and Berendzen, J. (1999) Automated MAD and MIR structure solution. *Acta Crystallogr D Biol Crystallogr*, **55**, 849-861.
- Theissen, G. and Melzer, R. (2007) Molecular mechanisms underlying origin and diversification of the angiosperm flower. *Ann Bot (London)*, **100**, 603-619.
- Underhill, D.A. (2000) Genetic and biochemical diversity in the Pax gene family. *Biochem Cell Biol*, **78**, 629-638.
- van Pouderooyen, G., Ketting, R.F., Perrakis, A., Plasterk, R.H. and Sixma, T.K. (1997) Crystal structure of the specific DNA-binding domain of Tc3 transposase of *C.elegans* in complex with transposon DNA. *Embo J*, **16**, 6044-6054.
- Weigel, D., Alvarez, J., Smyth, D.R., Yanofsky, M.F. and Meyerowitz, E.M. (1992) *LEAFY* controls floral meristem identity in Arabidopsis. *Cell*, **69**, 843-859.
- Weigel, D. and Jackle, H. (1990) The fork head domain: a novel DNA binding motif of eukaryotic transcription factors? *Cell*, **63**, 455-456.
- Wu, X., Dinneny, J.R., Crawford, K.M., Rhee, Y., Citovsky, V., Zambryski, P.C. and Weigel, D. (2003) Modes of intercellular transcription factor movement in the *Arabidopsis* apex. *Development*, **130**, 3735-3745.
- Xu, H.E., Rould, M.A., Xu, W., Epstein, J.A., Maas, R.L. and Pabo, C.O. (1999) Crystal structure of the human Pax6 paired domain-DNA complex reveals specific roles for the linker region and carboxy-terminal subdomain in DNA binding. *Genes Dev*, **13**, 1263-1275.

Figure legends

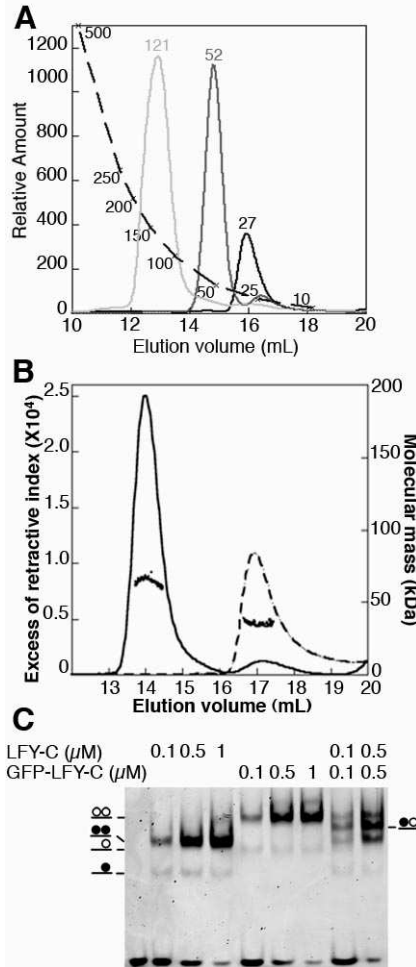


Figure 1. DNA dependent dimerization of LFY-C.

(A) Size exclusion chromatography. LFY-C (40 μ M, black curve), AP1 DNA (10 μ M, dark grey curve), LFY-C (40 μ M) + AP1 DNA (10 μ M, light grey curve) were analyzed. LFY-C elution at a volume corresponding to 28 kDa is consistent with the monomer size (25.7 kDa), the DNA duplex elutes earlier than expected at a volume corresponding to 52 kDa because of its elongated shape. The LFY-DNA complex elutes at a volume corresponding to 121 kDa. Molecular weights estimated from the calibration curve (dashed line) are indicated.

(B) Molecular mass of LFY-C alone (dashed line) or in combination with AP1 DNA (solid line) determined by multi-angle laser light scattering and refractometry combined with size exclusion chromatography. Elution profiles were monitored by excess refractive index (left ordinate axis). Dots show the molecular mass distribution (right ordinate axis). Average molecular mass is 64 ± 2 kDa for the LFY-C/DNA complex (65 kDa theoretical size for a dimeric complex) and 35 ± 1 kDa for LFY-C alone (26 kDa theoretical size for LFY-C monomer).

(C) Electromobility shift assay (EMSA) with 10 nM AP1 DNA and various LFY-C or GFP-LFY-C concentrations. Schematic complexes with LFY-C (filled circle) and GFP-LFY-C (open circle) are depicted.

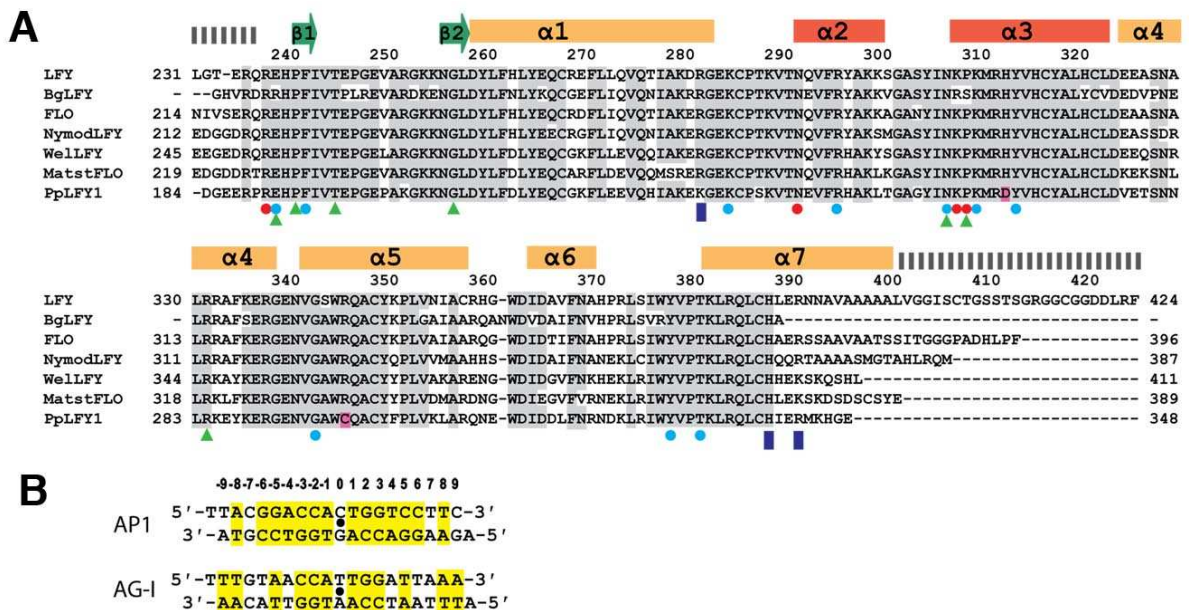


Figure 2: Sequence alignments

(A) Aligned carboxy-terminal amino acid sequences of LFY (*Arabidopsis thaliana*, AAA32826), BgLFY (*Brownea grandiceps*, AAS79888), FLO (*Antirrhinum majus*, P23915), NymodLFY (*Nymphaea odorata*, AAF77609), WellLFY (*Welwitschia mirabilis*, AAF23870), MatstLFY (*Matteuccia struthiopteris*, AAF77609), and PpLFY1 (*Paspalum paspalodes*, AAF77609).

AAF77608) and PpLFY1 (*Physcomitrella patens*, BAD91043). Identical and conservatively substituted residues are depicted on a grey background. Secondary structure elements are indicated. Residues involved in interactions with DNA bases and backbone are labeled with red and blue circles, respectively. Dashed bars indicate disordered regions in the crystal, blue rectangles the residues involved in dimerization. Green triangles indicate the position of *Arabidopsis* mutations, residues divergent in PpLFY1 are highlighted in pink.

(B) Two DNA duplexes containing the LEAFY binding sites from *AP1* and *AG* promoters present in the LEAFY/DNA complex crystals are depicted. Base pairs related by a dyad (indicated by a black dot) are highlighted in yellow.

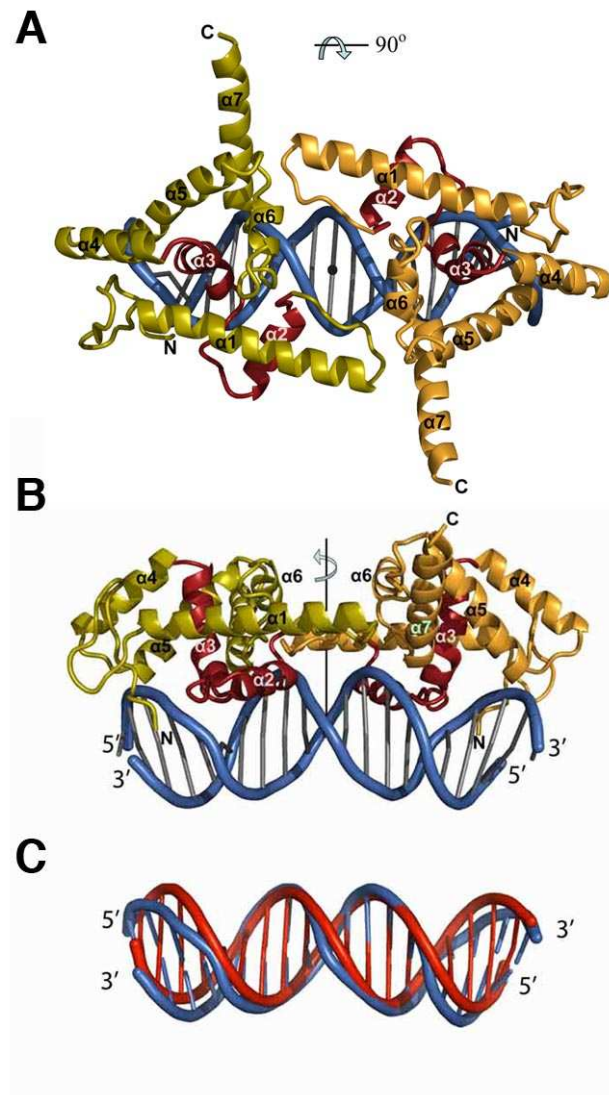


Figure 3: Structure of the LFY-C dimer bound to DNA

(A-B) Two orthogonal views of the LFY-C dimer (residues 237-399) bound to DNA. Monomers are colored in olive and orange with the helix-turn-helix motif (HTH, helices $\alpha 2$ and $\alpha 3$) in red. The DNA duplex is depicted in blue. Figures 3, 4B, 5A and 6 were produced with program Pymol (Delano, 2002).

(C) Superposition of the DNA duplex found in the LEAFY-DNA complex (blue) with regular B-form DNA (red).

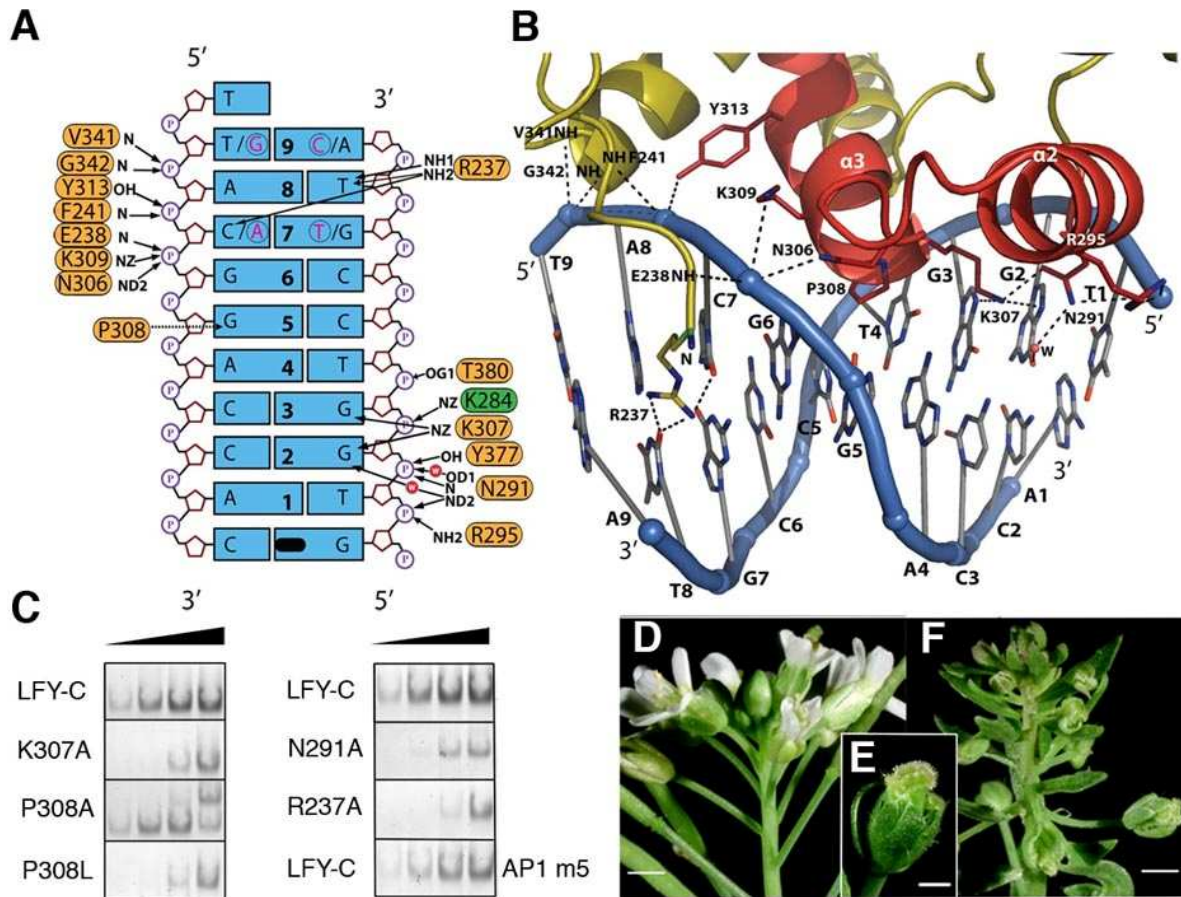


Figure 4: DNA recognition by LEAFY

(A) Protein-DNA interactions in one *AP1* half-site. Dyad-related base pairs 7 and 9 from the other half site are shown in pink and encircled. Polar and hydrophobic interactions are shown with solid and dashed arrows respectively. K284 belongs to the other monomer and is depicted in green. The pseudo-dyad coinciding with the crystallographic dyad is depicted in black.

(B) Ribbon diagram of one LEAFY monomer bound to its *AP1* half site. The protein is colored in olive except for the HTH motif shown in red. Polar interactions are indicated by dashed lines. For clarity only side chains in contact with DNA are shown.

(C) Effect of selected mutations on LFY-C DNA binding affinity to *AP1* DNA. EMSA were performed with wild-type and mutant LFY-C (100-250-750-2000 nM from left to right). Only dimeric complexes are shown except for P308A that gave rise to an unknown higher complex. *AP1* m5 mutant DNA contains base pair C:G instead of A:T at position ± 8 (see table S3 for full DNA sequences).

(D-F) Phenotype of the wild-type *Arabidopsis* inflorescence (D) and *lfy-28* (P308L) mutant inflorescence (F) and flower (E). Scale bar is 1 mm on D, F and 0.5 mm on E.

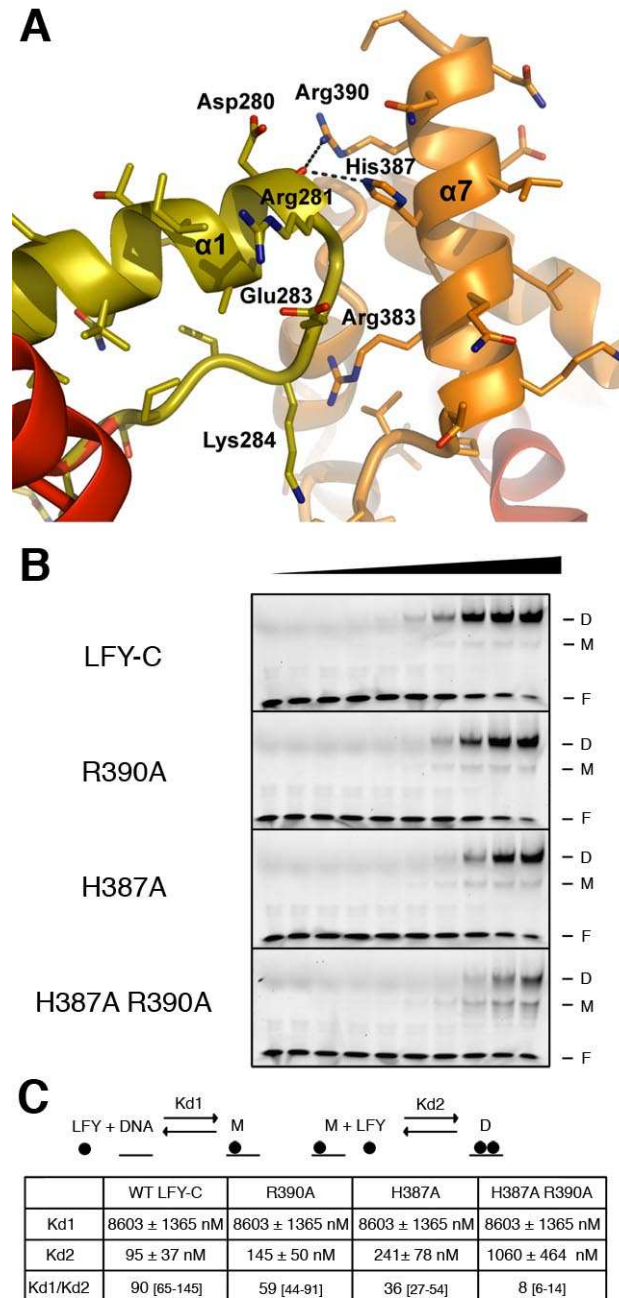


Figure 5: The LFY-C dimer interface mediates cooperative binding

(A) The dimer interface is viewed perpendicular to the DNA axis. Polar contacts between the two monomers (in orange and olive) are shown with dashed lines.

(B) EMSA with increasing concentrations (0, 10, 20, 50, 100, 200, 500, 1000, 2000, 3000 nM from left to right) of LFY-C wild-type, R390A mutant, H387A mutant, and H387A/R390A double mutant and 50 μ M AP1 DNA. Free DNA (F), monomeric (M) and dimeric (D) complexes are indicated.

(C) Estimation of dissociation constants for wild-type LFY-C and three mutant versions (H387A, R390A and H387A/R390A). Binding of LFY-C to AP1 DNA was modeled as two equilibrium reactions as detailed in Supplementary data: 1) Binding of a first LFY-C monomer to AP1 DNA, leading to the formation of the monomeric complex (M) and characterized by the Kd1 dissociation constant. 2) Binding of a second LFY-C monomer to M, leading to the formation of the dimeric complex (D) and characterized by Kd2. EMSA signals from B were quantified and the corresponding experimental values were fitted with theoretical equations describing the two equilibria. The errors and intervals between square brackets indicated correspond to the 95 % confidence interval. An elevated Kd1/Kd2 ratio reflects a high level of cooperativity whereas a ratio of 1 would indicate an absence of binding cooperativity. The single mutations resulted in a weak decrease of cooperativity whereas the H387A/R390A double mutation strongly decreased the cooperativity.

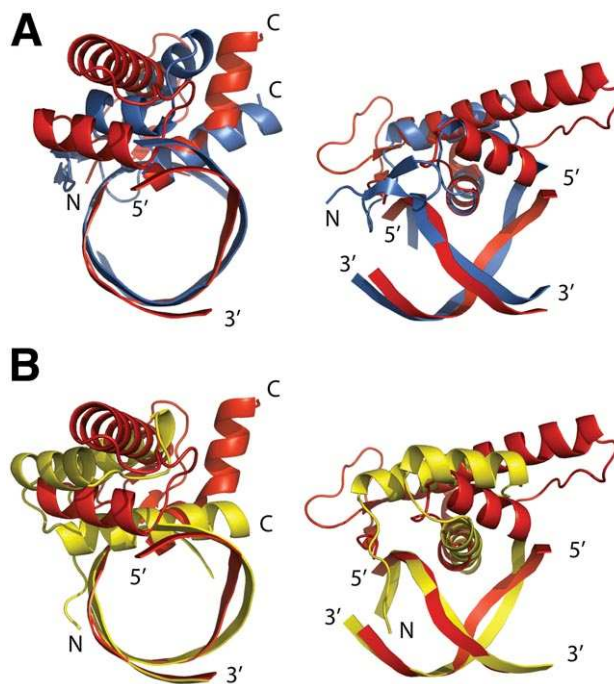


Figure 6: Comparison of LFY-C with paired and homeodomain DNA-binding

(A) Two orthogonal views of LFY-C helices $\alpha 1$ to $\alpha 3$ bound to their DNA target site (red) superimposed with the 3-helical bundle core of the N-terminal subdomain of the paired domain of *Drosophila* Prd (blue, PDB-id: 1pdn).

(B) Superposition with the homeodomain of *Drosophila* engrailed bound to DNA (yellow, PDB-id: 1hdd), where the center of recognition helix $\alpha 3$ inserts into the major groove.

Table 1. Structure determination of the LEAFY-DNA complex

Data Statistics						
Data set	Resolution (Å) ^a	Reflections measured/unique	R _{meas} ^b	I/σ	Completeness (%)	
AP1/LFY: space group P6 ₅ 22, unit cell dimensions a = b= 98.8 Å, c = 177.4 Å						
Native	20-2.1 (2.2-2.1)	503443/ 29859	4.8% (54.3%)	38.5 (5.0)	97.4 (83.9)	
EMTS	20-2.4 (2.5-2.4)	283416/ 20463	11.1% (50.2%)	15.5 (4.7)	99.6 (99.3)	
AG-I/LFY: space group P6 ₅ 22, unit cell dimensions a =b= 98.4 Å, c = 176.4 Å						
Native	20-2.3 (2.4-2.3)	328932/ 23456	7.9% (75.1%)	26.7 (3.4)	99.8 (100.0)	

Phasing statistics for the EMTS derivative (SIRAS)	
Wavelength (Å)	0.934
Phasing power ^c	1.68
Figure of merit	0.468
R _{cullis} ^d	0.674
Number of mercury sites	4

Refinement statistics		
	LFY/AP1	LFY/AG-I
Resolution ^a	19.1 - 2.1 (2.2 - 2.1)	20.0 - 2.3 (2.4-2.3)
Total number of non-hydrogen protein atoms	1352	1332
Total number of non-hydrogen DNA atoms	526	567
Number of water molecules	148	105
R-factor (%) ^a	21.0 (25.0) for 26896 reflections	22.1 (29.0) for 22559 reflections
R _{free} (%) ^{a, e}	23.7 (26.4) for 1521 reflections	24.9 (30.7) for 1183 reflections
R.m.s. deviations:		
Bond lengths (Å)	0.009	0.008
Bond angles (°)	1.22	1.21
Average Temperature factors (Å ²):		
Protein	41.1	39.8
DNA	48.1	46.4
Solvent	42.8	48.5
R.m.s. deviations of covalently linked atoms (Å ²)	2.55	3.45
Residues in Ramachandran plot: ^f		
Most favored regions (%)	93.2 (136)	95.2 (139)
Additionally allowed regions (%)	6.8 (10)	4.8 (7)
Generously allowed regions (%)	0 (0)	0 (0)
Disallowed region (%)	0 (0)	0 (0)

a Values for the highest resolution range are given in parenthesis.

b R_{meas} is a redundancy independent R-factor as defined in (Diedrichs & Karplus, 1997).

c Phasing power is the mean value of the heavy-atom structure factor amplitudes divided by the lack of closure.

d R_{cullis} is the mean lack of closure divided by the mean isomorphous difference.

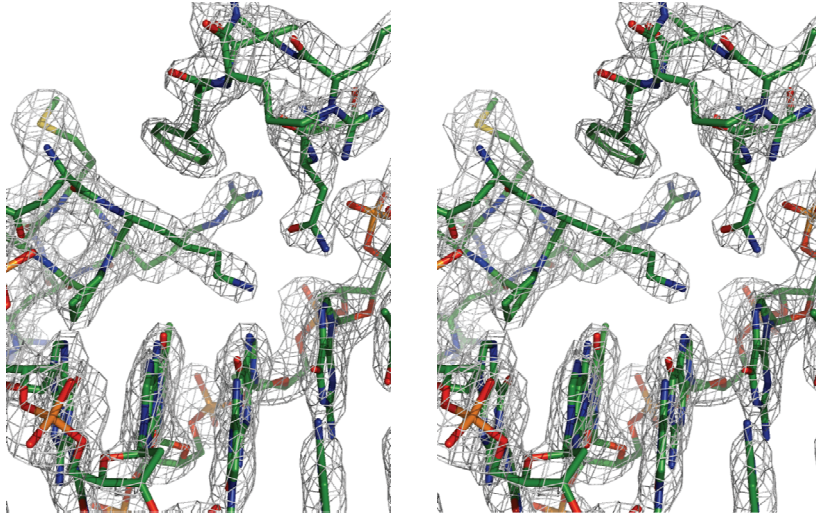
e R_{free} was calculated from a subset of 5% of the data.

f Number of residues are given in parentheses.

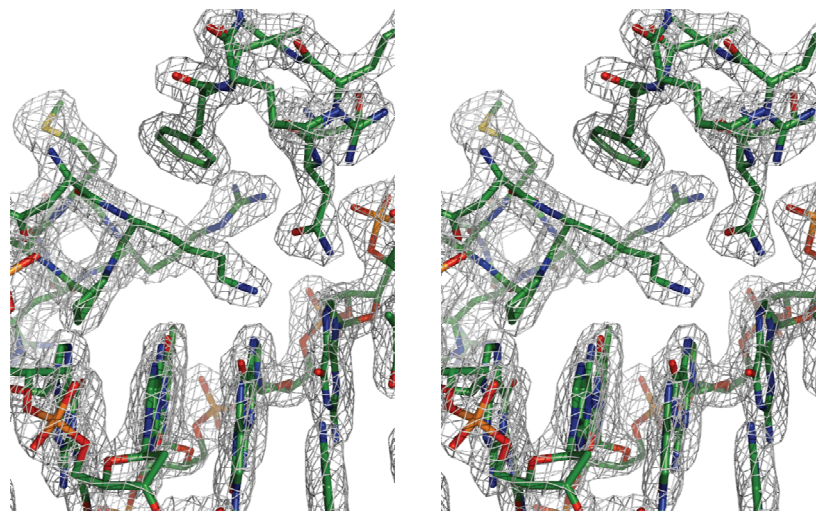
Supplementary Figures

Supplementary Figure 1: Comparison of the initial experimental SIRAS electron density map after solvent flattening (A) and the final 2Fo-Fc electron density map (B). Both maps are contoured at 1.9 σ .

A

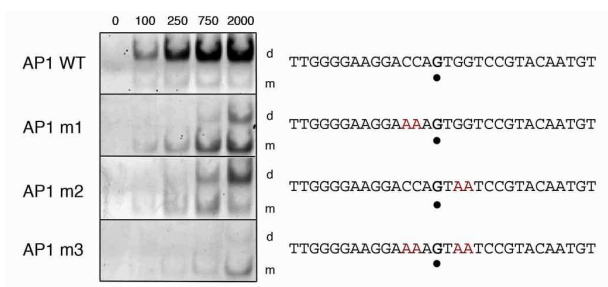


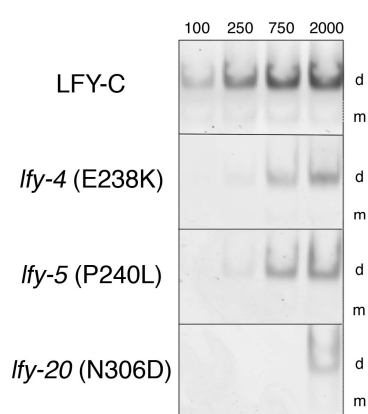
B



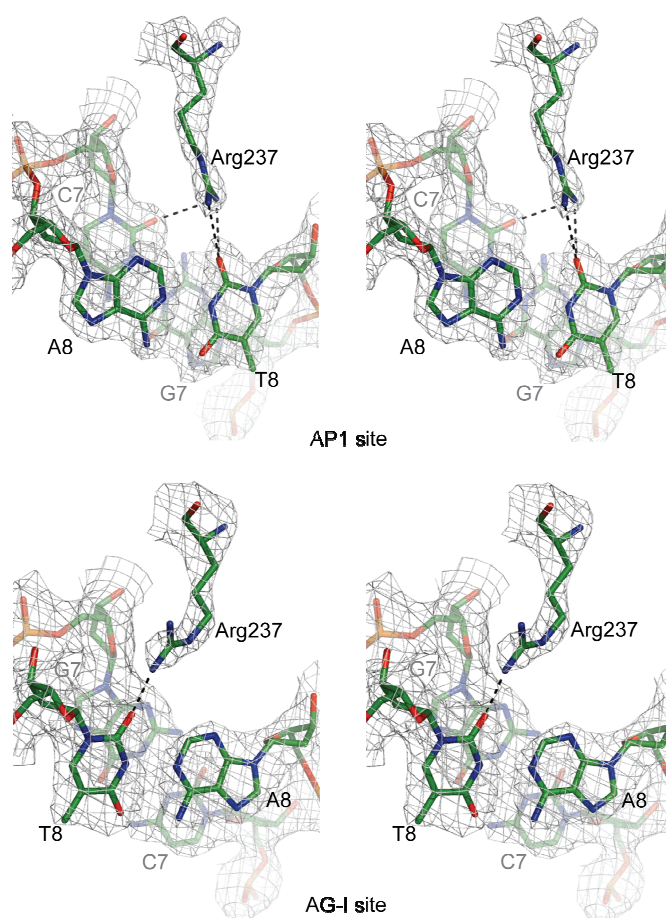
Supplementary Figure 2:

EMSA with wild-type and mutant *AP1* probes (10 nM) and wild-type LFY-C (concentrations in nM). Mutations in binding sites resulted in a reduction of both binding affinity and cooperativity as shown by the increased monomeric complex (m) concentration and reduction of the dimeric one (d). 5' to 3' sequence of the DNA upper strand is indicated on the right of the EMSA. The dots indicate the center of the pseudo-palindromic binding site.

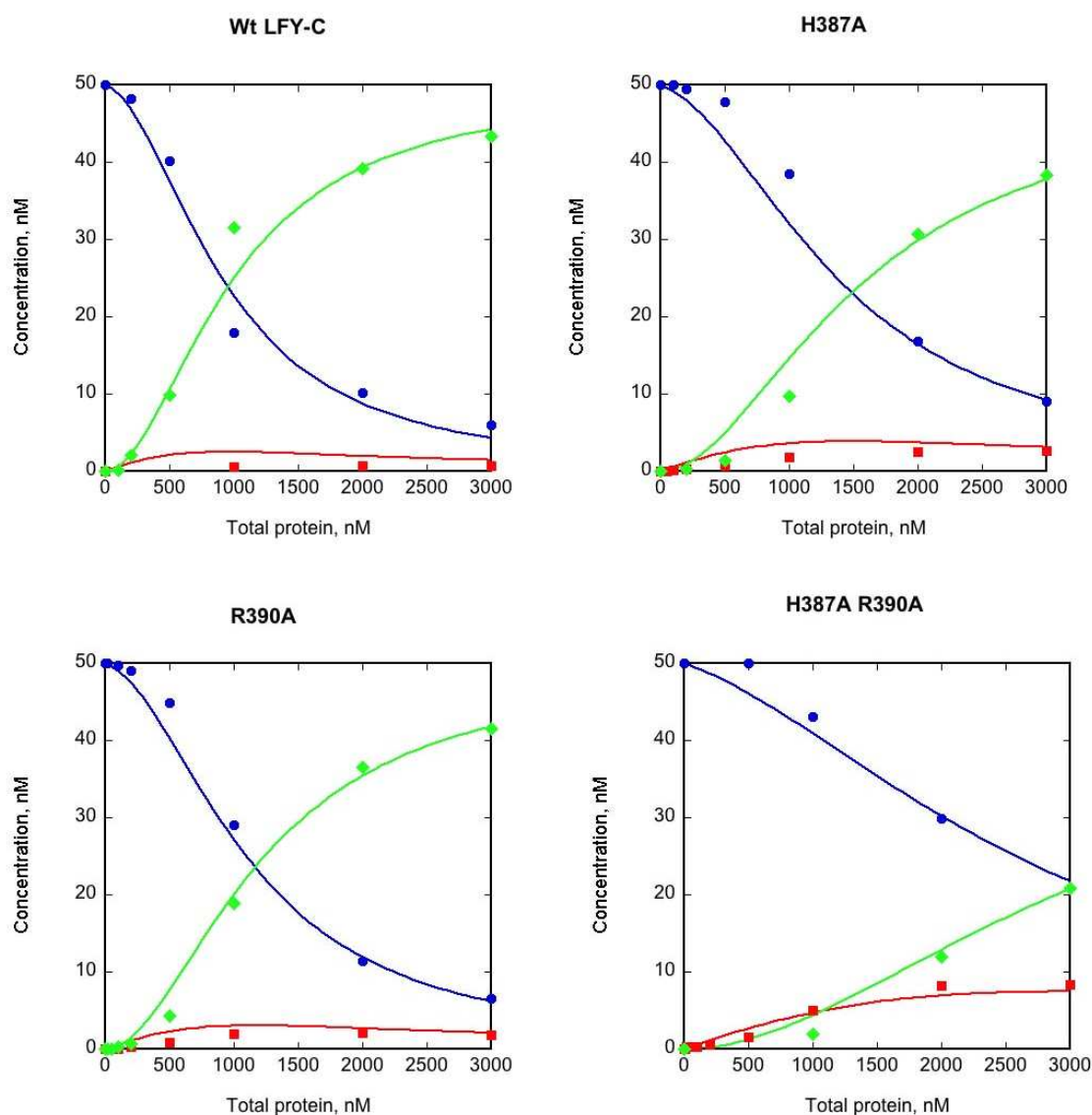




Supplementary Figure 3: EMSA with wild-type and mutant LFY-C proteins (concentrations in nM). *AP1* DNA concentration is 10nM. Positions of dimeric (d) and monomeric (m) complexes are indicated. These mutations had been found to abolish DNA binding in previous experiments, probably because the binding was only tested under conditions where the wild-type protein shifts only small amounts of the DNA probe



Supplementary Figure 4: Stereo diagram of the interactions between R237 and base pairs 7 and 8 observed in the *AP1* (top) and *AG-I* (bottom) sites. Depicted are the final 2Fo-Fc electron density maps at 1.4 Å. The side chain of Arg237 is in contact distance to base pairs deviating from 2-fold symmetry (bp₊₉ and ₊₇ in *AP1*; bp_{±7} in *AG-I*). In the *AP1* site Arg237 NH1 and NH2 interact with the exocyclic O2 of Thy₊₈ in the *AP1* site. In one halfsite Arg237 NH1 contacts the O2 of Cyt₊₉ of the same strand, while in the other halfsite Arg237 NH2 contacts O2 of Cyt-7 of the opposite strand (depicted here). Arg237 possesses well-defined side chain density indicating a unique conformation despite the differences in the two halfsites. In the *AG-I* site, Arg237 NH1 contacts the O2 of Thy₊₈, but is too far away to form direct hydrogen bonds with the bases in base pair ₊₇. In the *AG-I* site the electron density for the Arg237 side chain is slightly less well defined as in the *AP1* site, but we do not find evidences for alternative conformations of this side chain.



Supplementary Figure 5: Quantitative analysis of cooperativity

Quantifications of free DNA (blue), monomeric complex M (red) and dimeric complex D (green) were obtained from experimental data shown in Figure 5 and plotted as circles, squares and diamonds, respectively. Calculated fits are represented as plain lines following the same color code.

Supplementary Table 1: Sequences of oligonucleotides used for EMSA. Only forward sequences are indicated (5' to 3'). Mutant bases are underlined. The guanine corresponding to the center of the pseudo-palindromic binding site is depicted in bold.

Oligonucleotide	Sequence
AP1 WT	TTGGGGAAGGACCA G TGGTCCGTACAATGT
AP1 m1	TTGGGGAAGGAAA A GTGGTCCGTACAATGT
AP1 m2	TTGGGGAAGGACCA G T A ATCCGTACAATGT
AP1 m3	TTGGGGAAGGAAA A GT A ATCCGTACAATGT
AP1 m5	TTGGGGCAGGACCA G TGGTCCG G ACAATGT

Supplementary Methods

Determination of dissociation constants.

Signals were quantified from the gels shown in Figure 5 using ImageQuant software (Molecular dynamics, Sunnyvale, CA). The sum of the signals per lane was normalized to 100% and the concentrations of free DNA (A), monomeric (M) and dimeric (D) complexes were calculated as a fraction of total fluorescent DNA concentration (50 nM in these experiments). The fits were performed assuming K_{d1} is the same for all 4 proteins (LFY-C, H387A, R390A and H387A R390A). Without this assumption, K_{d1} values were found to be very similar to each other but with larger confidence intervals.

Model variables and parameters.

Variable	Symbol	Unit
Total DNA	A_0	nM
Total protein	C_0	nM
Free DNA	A	nM
Free protein	C	nM
Protein.DNA	M	nM
(Protein) ₂ .DNA	D	nM
First dissociation constant	K_{d1}	nM
Second dissociation constant	K_{d2}	nM

Chemical reaction	Mass-action constraints
$A + C \rightleftharpoons M$	$K_{d1} = \frac{[A][C]}{M}$
$C + M \rightleftharpoons D$	$K_{d2} = \frac{[M][C]}{D}$

Conservation relation	
Total DNA	$A_0 = [A] + [M] + [D]$
Total protein	$C_0 = [C] + [M] + 2 \cdot [D]$

Mathematical analysis and least-square solution.

From the mass-action conditions, we get the concentration of the two complexes M and D as a function of free DNA and protein concentrations:

$$[M] = \frac{[A][C]}{Kd_1}$$

$$[D] = \frac{[M][C]}{Kd_2} = \frac{[A][C]^2}{Kd_2 \cdot Kd_1}$$

The DNA conservation condition together with the two above relations, gives free DNA as function of free protein.

$$[A] = \frac{A_0}{1 + \frac{[C]}{Kd_1} + \frac{[C]^2}{Kd_2 \cdot Kd_1}} = \frac{Kd_2 \cdot Kd_1 \cdot A_0}{Kd_2 \cdot Kd_1 + Kd_2 \cdot [C] + [C]^2} \quad (1)$$

Finally, using relation (1) and the total protein concentration, we obtain a third-degree polynomial expression for the unknown variable [C]:

$$P([C]) = -C_0 \cdot Kd_2 \cdot Kd_1 + (Kd_2 \cdot Kd_1 - C_0 \cdot Kd_2 + A_0 \cdot Kd_2) \cdot [C] + (2A_0 + Kd_2 - C_0) \cdot [C]^2 + [C]^3 \quad (2)$$

It is easy to check that $P(0) = -C_0 \cdot Kd_1 \cdot Kd_2 < 0$ and that $P(\infty) = 1$; hence, it exists at least one positive real root of the equation $P([C]) = 0$ for $[C] > 0$. Let c be this root. Once this value is known, the equilibrium concentrations of the other molecular species are directly obtained as:

$$[A] = \frac{Kd_1 \cdot Kd_2 \cdot A_0}{Kd_1 \cdot Kd_2 + Kd_2 \cdot c + c^2},$$

$$[C] = c,$$

$$[M] = \frac{[A] \cdot c}{Kd_1}, \quad (3)$$

$$[D] = \frac{[A] \cdot c^2}{Kd_1 \cdot Kd_2}.$$

The unknown parameters of the global problem are the Kd_1 dissociation constant, which is independent of the protein interacting with the DNA molecule, and four dissociation constants, denoted $Kd_{2,LFY-C}$, $Kd_{2,H387A}$, $Kd_{2,R390A}$ and $Kd_{2,H387A R390A}$. For each set of experimental data, we derive a distance between the concentration of free DNA, M and D, as predicted by solving eqs. 2-3, and the experimental counterpart:

$$E_X = \sum_{i=1, \dots, N} ([A](C_{X,0,i}) - [A]_{exp,i})^2 + \sum_{i=1, \dots, N} ([M](C_{X,0,i}) - [M]_{exp,i})^2 + \sum_{i=1, \dots, N} ([D](C_{X,0,i}) - [D]_{exp,i})^2,$$

where X represents one of the proteins under investigation (i.e. LFY-C, H387A, R390A and H387A R390A); $[A]_X$, $[M]_X$ and $[D]_X$ are given by eq.3; $[A]_{exp}$, $[M]_{exp}$ and $[D]_{exp}$ are the experimental measures of the free DNA and the two DNA-protein complexes concentrations. Note that the parameter $C_{X,0,i}$, which codes for the i th total protein concentration (for $X = LFY-C$, H387A, R390A and H387A R390A) enters the model through equation (2). Because we consider all experimental data globally, we determine the best parameter set by minimizing the distance between model and experiments:

$$E = E_{LFY-C} + E_{H387A} + E_{R390A} + E_{H387AR390A}.$$

Then, once the best parameter set is determined, the second order variation of the distance E close to the best fit gives confidence interval for the different parameters and their ratio.

Annexe 2

Article à soumettre

The Analysis of Entire Gene Promoters by Surface Plasmon Resonance

Edwige Moyroud^{1,2#}, Mathieu C. A. Reymond^{1#}, Cécile Hames², François Parcy^{2*} and Charles P. Scutt^{1*}

Author addresses

¹Laboratoire de Reproduction et Développement des Plantes,
(UMR CNRS 5667 - INRA - ENSL - UCBL),
Université de Lyon, 46 allée d'Italie,
69364 Lyon Cedex 07, France

²Laboratoire de Physiologie Cellulaire Végétale,
(UMR CNRS 5168 - CEA - INRA - UJF),
Bâtiment C2, 17 rue des Martyrs,
38054 Grenoble Cedex 9, France

#These two authors contributed equally to this work.

*Joint authors for correspondence.

Abstract

We demonstrate that the biophysical technique of Surface Plasmon Resonance (SPR), which has previously been used to measure transcription factor binding to short DNA molecules, can also be used to characterise transcriptional interactions involving entire promoters measuring several kilobases. The consequences of this discovery are considerable. Firstly, SPR can be used as a simple qualitative test of whether a given gene may, or may not, be direct target of a transcription factor of interest. This application will be particularly useful for the confirmation of predicted transcriptional interactions in model species, and for comparative studies of transcriptional interactions in non-model species that are not amenable to study by other methods. Secondly, SPR analysis can be used quantitatively to define the parameters that determine the binding affinity of transcription factor complexes to multiple sites in entire gene promoters. In this way, SPR could be used to test models for the complex combinatorial action of transcription factors that cannot be tested using existing methods. These advantages combine to make the SPR analysis of entire promoters a uniquely powerful technique, which should be of particular application in systems biology, in evolutionary-developmental biology, and in various branches of applied biology.

Introduction

Transcriptional regulation lies at the heart of most biological processes throughout the living world. This form of regulation depends on the physical interactions that take place between transcription factors and short DNA binding sites present in the cis-regulatory regions (promoters etc) of their target genes. However, the level of transcription of a given gene may depend on the positions, numbers and affinities of the cis-acting sites within its promoter and it would therefore be more biologically relevant to measure transcription factor binding to entire promoters than to individual cis-acting sites. Entire promoters can be analysed using *in vivo* techniques such as reporter gene assays. However, the use of *in vivo* methods is mainly restricted to model species, and these techniques do not elucidate the biophysical processes taking place at the level of the gene promoter. By contrast, a biophysical method, capable of analysing transcription factor binding to whole promoters, would be available for use in any organism, and would provide detailed biophysical data on the interaction under study.

Surface Plasmon Resonance has been used to quantify many types of molecular interactions, including the binding of transcription factors to short DNA molecules. In this method, interactions are typically measured between DNA attached to the surface of a gold-coated chip and solutions of transcription factors flowing over the chip. Molecular interactions are measured from the angular deflection of a band of extinction that occurs within a beam of plane-polarised light reflected from the chip's surface. Under typical experimental conditions, the coupling of this optical deflection to molecular interactions extends for approximately 200 nm into the solution (1,2). We reasoned that this range should be sufficient to measure interactions involving immobilized DNA molecules of considerable length, given that the flow of solution, and the non-linear average conformation of the DNA molecules, would tend to bring interaction sites closer to the chip's surface.

We present here a proof of concept for the use of SPR analysis to measure transcriptional interactions involving long DNA molecules, the size of typical eukaryotic gene promoters. We show this technique to be capable of both qualitative use, to discriminate between target and non-target genes of a given transcription factor, and quantitative use, to investigate the importance to transcriptional regulation of changes in overall DNA-binding affinity between transcription factors and entire promoters. SPR analysis can be performed using recombinant proteins and promoters from any species, and will thus enable the study of transcriptional interactions in non-model organisms. We discuss the potential qualitative and quantitative uses of this novel application of SPR in various fields of biology in which transcriptional regulation is of key importance.

Material and Methods

Expression and purification of recombinant protein.

The DNA binding domain of the LEAFY transcription factor from *Arabidopsis thaliana* (LFY-DBD), corresponding to amino acids 223 to 424 of the full-length protein, was inserted into the *pETM11* expression vector (3) to yield a plasmid, *pCH28*, permitting the production of LFY-DBD fused to an N-terminal 6x HIS Tag. Production of this protein was initiated by the addition of IPTG, to a final concentration of 0.5 mM, to cultures of *pCH28*-harboring *E. coli* cells of the Rosetta Blue DE3-pLysS strain (Novagen, Strasbourg, France). Following incubation overnight at 22°C, the cell cultures were lysed by sonication, and tagged, recombinant LFY-DBD protein was purified by affinity chromatography on Ni-NTA resin (Qiagen, Courtaboeuf, France). Protein-containing fractions were pooled and subjected to size-exclusion chromatography using a Hi-load Superdex-200 16/60 preparation grade column (GE Healthcare, Orsay, France) to eliminate protein aggregates. Soluble protein concentration was estimated using the Bradford assay (4).

Preparation of biotinylated DNA molecules for SPR analysis

DNA fragments containing cis-regulatory regions were obtained by PCR amplification, using the oligonucleotide primers (Eurogentec, Belgium) shown in Table 1, as follows: the 5'-flanking region (promoter) of *APETALA3* (*AP3*) and the second intron of *AGAMOUS* (*AG*) were amplified from genomic DNA of the Columbia ecotype of *Arabidopsis thaliana*; the promoter of *AP1*, also of the Columbia ecotype, was amplified from a plasmid provided by Dr Reyes Benlloch (CNRS, Grenoble); and the promoter and coding sequence of the heat shock protein 70 (*hsp70*) gene was amplified from human genomic DNA. The DNA fragments obtained were ligated into the pGEM-T Easy vector (Promega, Madison, WI). A *pBluescript*-derived plasmid containing part of the promoter and coding sequence of the BHLF1 gene, corresponding to bases 52401 to 53092 of the Epstein-Barr virus genome (Genbank accession V01555), was obtained from Dr Henri Gruffat (ENS-Lyon). Mutant versions of the second intron of *AG*, containing mutations to disrupt the function of LFY-binding sites as shown in Figure 1, were constructed by sequential site-directed mutagenesis (5). Double-stranded DNA molecules for SPR analysis were synthesized from recombinant plasmids by PCR amplification using M13-forward and reverse sequencing primers, one of which was biotinylated in each case, and subsequently purified using a NucleoSpin® Extract II kit (Macherey-Nagel, Düren, Germany).

Immobilization of DNA samples for SPR analysis.

CM5 SPR chips (Biacore, Uppsala, Sweden) were activated in a Biacore T100 SPR Analyser to accept a streptavidin coating using an amine coupling kit (Biacore) according to the manufacturer's instructions. Amine-coupling reagents were injected in a continuous flow of 5 µl/min of HBS-P+ buffer (0.01 M HEPES pH 7.4, 0.15 M NaCl, 0.05% w/v P20 detergent). Aliquots of 50 µl of streptavidin (0.1 mg/ml in 10 mM sodium acetate pH 4.2) were then injected, followed by 50 µl aliquots of ethanolamine (1M, pH 8.5) to inactivate residual carboxyl groups. Chips were then washed by two injections of 5 µl of HBS-P+ buffer. Double-stranded DNA molecules (5 ng/µl in HBS-P+ buffer, degassed before use) each carrying a single biotin moiety at its 3'-terminus (relative to the direction of transcription of the encoded gene), were immobilized in separate

channels on SPR chips by injection at 10 $\mu\text{l}/\text{min}$, until 200 RU (arbitrary SPR units) of DNA had been added, corresponding to approximately 0.15 ng/mm². One channel was left blank on each chip as a reference. Unbound DNA was finally removed by injections of 10 μl aliquots of NaCl (1 M).

SPR analysis of DNA-protein interactions

SPR chips containing immobilized DNA samples were equilibrated in a Biacore T100 Analyser by injection of HBS-P+ buffer containing non-homologous DNA (40 ng/ μl , molecular size >120 bp) from salmon testis (Roche, Meylan, France) until SPR responses were stable. Transcription factors, dissolved in the above buffer, were then injected into all channels of these chips for 250 s at a flow rate of 50 $\mu\text{l}/\text{min}$, followed by injection of further buffer for 300 s to monitor protein dissociation. Transcription factor solutions were used in decreasing order of concentration and SPR chips were regenerated between analyses by the sequential injection of guanidinium hydrochloride (3 M) and sodium dodecyl sulfate (0.03%, w/v), each for 60 s at a flow rate of 50 $\mu\text{l}/\text{min}$.

SPR data analysis

Real-time SPR interaction curves were analyzed using BioEval T100 software (Biacore, Uppsala, Sweden). For each analysis, the response of the reference channel was subtracted from the interaction curves obtained from the three experimental channels and the resulting curves were then adjusted to zero at the start of transcription factor injection. These normalized curves were then fitted globally to a 1:1 (Langmuir) interaction model, permitting the determination of forward and reverse apparent rate constants ($k_{\text{on}}^{\text{APP}}$ and $k_{\text{off}}^{\text{APP}}$). The validity of the interaction model was verified by data-fitting, a good fit indicated by $\chi < 2$ (or < 10 for low affinity interactions). Equilibrium binding constants ($K_{\text{D}}^{\text{APP}}$) were calculated from the ratio $k_{\text{off}}^{\text{APP}}/k_{\text{on}}^{\text{APP}}$ and also from SPR values (RU) at equilibrium over a range of protein concentrations.

Results

SPR analysis can be used to distinguish between target and non-target genes.

In many cases, it would be useful to determine, by a simple and rapid physical assay, whether a particular gene can, or cannot, be bound by a given transcription factor. To test the suitability of SPR for this purpose, we studied binding interactions involving the well-characterised transcription factor LEAFY (LFY) from the model plant *Arabidopsis thaliana*. LFY controls floral patterning by inducing the expression of genes encoding several members of the MADS box family of transcription factors (6,7). We compared the binding of the LFY DNA-binding domain (LFY-DBD) (8) to the promoters of two of its MADS-box target genes, *AP1* and *AP3*, and to two negative control DNAs derived from the promoters of the Epstein-Barr virus BHLF1 and human hsp70 genes. The *AP1* and *AP3* DNA molecules used in these assays measured 2.2 kb and 2.9 kb, respectively, and both contained at least one experimentally confirmed LFY-binding site (7,9). Equal quantities, in arbitrary SPR response units (RU), of biotinylated DNA samples were immobilised in separate channels on streptavidin-coated SPR chips. Interactions between LFY-DBD and immobilized DNA molecules were then monitored by SPR over a range of protein concentrations (Fig. 2a-c). These analyses were performed in the constant presence in solution of non-specific competitor DNA to avoid signal saturation by non-specific binding.

Differences between target and non-target genes of LFY were visually apparent in the SPR interaction curves obtained: low concentrations of LFY-DBD gave higher SPR responses to target (Fig 2a and b) than to non-target (Fig 2c) DNA, while LFY-DBD dissociated rapidly and completely from non-target DNA, but slowly and incompletely from target DNA. $K_{\text{D}}^{\text{APP}}$ values were calculated from real-time interaction data using dedicated SPR analysis software (Tab 2), assuming a 1:1 (Langmuir) binding model. This method of calculation also provided a statistical test to verify the appropriateness of this binding model. As an independent verification, $K_{\text{D}}^{\text{APP}}$ values were also calculated from the linear gradients of graphs based on the equilibrium condition (Fig 2d). The linearity of these plots provided an independent verification of the appropriateness of using a 1:1 binding model. $K_{\text{D}}^{\text{APP}}$ values obtained using both methods of calculation were in good agreement, and fell within the 10-100 nM range for both the *AP1* and *AP3* target genes (Fig 2d and Tab 2). The measurement of interactions with non-target DNA of the BHLF1 and hsp70 genes yielded $K_{\text{D}}^{\text{APP}}$ values in excess of 1 μM , at least some two orders of magnitude higher than those for target genes. In the case of LFY, SPR analysis thus proved capable of easily distinguishing between target and non-target DNA measuring up to 3 kb in length.

To test the ability of SPR to discriminate between target and non-target genes using a second type of transcription factor, we analyzed interactions of the human heat shock transcription factor, HSF1, with the promoter of its predicted target gene hsp70, and with a negative control DNA derived from the promoter of the Epstein-Barr virus gene, BHLF1. The hsp70 DNA fragment used in these assays measured 2.4 kb and contained one HSF1-binding site (10). The K_D^{App} for the interaction between HSF1 for hsp70 (Fig 3a), calculated both using dedicated SPR software (Tab 2) and from plots of equilibrium data (Fig 3b), fell within the nanomolar range. No interaction, however, was observed between HSF1 and non-target DNA from BHLF1 (Fig 3a). Hence, in the case of HSF1, SPR was able to completely discriminate between target and non-target DNA, even without quantitative analysis. Experiments using both LFY-DBD and HSF1 have thus shown SPR analysis to provide a simple and rapid means to test whether a gene promoter can, or cannot, be bound by a given transcription factor.

SPR provides quantitative data on transcription factor binding to entire cis-regulatory regions.

In many cases, the consequences of transcriptional interactions will depend quantitatively on the binding affinity of transcription factors for their target genes. SPR analysis is known to represent one of the most quantitative techniques available for the measurement of physical interactions between transcription factors and individual cis-acting binding sites. To test whether these quantitative characteristics are conserved when long DNA molecules are analyzed, we used SPR to investigate the binding of LFY to three cis-regulatory regions differing only in the number of cis-acting elements present. The second intron of AG is necessary for the correct transcriptional regulation of this gene by LFY. This intron measures 3.0 kb and contains four consensus LFY-binding sites (11). We made two mutated versions of the AG second intron to disrupt binding to two, or to all four, of its LFY-binding sites (Figure 3). K_D^{App} values for interactions between these variants of the AG second intron and LFY-DBD (Table 1) were calculated from real-time SPR interaction data on the basis of a 1:1 binding model, as described above. A K_D^{App} value of 0.6 nM was obtained for the wild-type intron, as compared to values of 42 nM and 270 nM for the doubly and quadruply mutated introns, respectively. K_D^{App} values for the overall interaction with the AG second intron thus increased progressively with the elimination of LFY-binding sites. These data clearly demonstrate SPR to retain a quantitative value when applied to long DNA fragments, a feature which will be of great importance for the detailed study of complex transcriptional interactions.

Discussion

Why length is so important.

Despite the availability of commercial SPR analyzers for the last 15 years or more, SPR analysis has not been widely taken up by researchers working on transcription factors. This lack of widespread interest has probably been due to the perception that SPR, as applied to transcriptional interactions, was limited to the analysis of short DNA fragments. Accordingly, this technique has mostly been used for detailed, quantitative studies of transcription factor binding to individual cis-acting sites, rather than to relate such biophysical events to their higher level effects. In the present work, we demonstrate two things: firstly that SPR can be used to detect transcription factor binding to much longer DNA fragments than was previously believed possible, and secondly that this technique retains its quantitative value when applied to long DNA molecules. The first of these findings means that SPR can be used as a simple and rapid, qualitative test of whether one gene may, or may not, be the direct target of another. The second finding opens the possibility of more subtle uses of SPR in the analysis of complex transcriptional interactions. Both the qualitative and quantitative uses of SPR should find many applications to biological problems, as described below.

To bind or not to bind: novel qualitative applications of SPR analysis.

A major objective of systems biology is to describe how networks of transcriptional regulators control complex biological processes. A first requirement of such studies is the identification of the direct target genes of transcription factors of interest, which can be achieved using such *in vivo* techniques as microarray analyses (12) and chromatin immunoprecipitation (13). These procedures yield lists of putative target genes, which must then be verified by independent analyses, including the *in vitro* analysis of DNA binding. Previously, such *in vitro* analyses have only been possible for transcription factors whose binding sites could be found within cis-regulatory regions. The SPR analysis of entire gene promoters, as demonstrated in the present work, will provide a rapid means of verifying such predicted transcriptional interactions, even where binding sites cannot be predicted from the DNA sequence. This is important not only for transcription factors whose binding site preferences are currently unknown, but also for the many cases in which transcription factors bind to biologically relevant, but cryptic, sites that show little similarity to their known consensus sequences. SPR machines of the latest generation are capable of analyzing several hundreds of bound DNA samples simultaneously, and should therefore prove ideal for the rapid verification of lists of direct target genes identified using large-scale approaches such as microarray analysis and chromatin immunoprecipitation. Indeed, with the possibility of analysing large numbers of bound DNA molecules simultaneously, SPR could also be used for the *in vitro* characterization of entire transcriptional networks by sequentially passing all transcription factors in a network over the complete set of that network's promoters. In this way, the transcriptional relationships linking all the components of a system, including hierarchical regulation, mutual regulation and auto-regulation, could be identified.

Differences in transcriptional control relationships between organisms probably account for much of the biodiversity of the natural world. However, our current knowledge of transcriptional interactions derives exclusively from the study of a few model organisms that are amenable to genetic analysis. To identify the differences in transcriptional interactions that are responsible for biodiversity, a means is needed to determine which interactions are conserved, and which are not conserved, between model and non-model organisms. The SPR analysis of entire gene promoters should prove ideal for this purpose as it can easily be performed using recombinant transcription factors and putative target genes from any species. The use of SPR to characterise transcriptional relationship in non-model species will be of great importance to both evolutionary-developmental (evo-devo) (14) and applied biology, in which the organisms of interest are often unsuitable to *in vivo* studies of transcriptional interactions.

Putting a figure on it: novel quantitative applications of SPR analysis.

In the present work, we have shown that the quantitative value of SPR analysis is conserved for interactions involving long DNA molecules. Accordingly, the SPR analysis of transcription factor binding to a given target promoter yields a single K_D^{App} value that quantifies the overall binding interaction between those two components. By repeating SPR analyses using mutated versions of a promoter of interest, it should be possible to determine the importance to the overall binding interaction of the positions, numbers and affinities of the binding sites present. Such an approach would rapidly indicate the presence of such phenomena as cooperative binding, where several sites of individually poor affinity may produce a strong overall binding interaction.

Transcription factors frequently bind to DNA as complexes, and these may also interact simultaneously with several cis-acting sites positioned at considerable distances along the target gene. For example, the MADS-box transcription factors encoded by three of the target genes analysed in the present work, *AP1*, *AP3* and *AG*, are hypothesized, on the basis of genetic evidence, to form various combinations of tetramers (15) with a further class of MADS-box protein, *SEPALLATA* (16,17). According to this hypothesis, four different tetrameric complexes of MADS box proteins would interact with multiple cis-acting sites in four different sets of target genes to specify the four types of organs in the flower: sepals, stamens, petals and carpels. No technique has previously been available to directly measure the affinity

of binding of these transcription factors, singly and in combination, to the entire cis-regulatory regions of their various sets of predicted target genes. However, SPR analysis should provide a means to test the combinatorial interactions of transcription factors with the entire promoters of their different sets of predicted target genes, and hence to test this and other complex models of transcriptional regulation.

Acknowledgements

We thank Professor Mike Frohlich for helpful discussions, Professor Sylvie Ricard-Blum and Clément Faye for advice on SPR analysis, and Drs Reyes Benlloch and Henri Gruffat for supplying plasmids. We acknowledge funding from the grants ANR-05-BLAN-0280-01 ("Carpelle Virtuel") to CS and ANR-07-BLAN-0211-01 ("Plant TF-Code") to FP and CS.

References

1. Lukosz, W. (1997) Integrated-optical and surface-plasmon sensors for direct affinity sensing .2. Anisotropy of adsorbed or bound protein adlayers. *Biosensors & Bioelectronics*, **12**, 175-184.
2. Kunz, R.E. and Cottier, K. (2006) Optimizing integrated optical chips for label-free (bio-)chemical sensing. *Analytical and Bioanalytical Chemistry*, **384**, 180-190.
3. Dummer, A., Lawrence, A.M. and de Marco, A. (2005) Simplified screening for the detection of soluble fusion constructs expressed in E-coli using a modular set of vectors. *Microbial Cell Factories*, **4**.
4. Bradford, M.M. (1976) Rapid and Sensitive Method for Quantitation of Microgram Quantities of Protein Utilizing Principle of Protein-Dye Binding. *Analytical Biochemistry*, **72**, 248-254.
5. Kirsch, R.D. and Joly, E. (1998) An improved PCR-mutagenesis strategy for two-site mutagenesis or sequence swapping between related genes. *Nucleic Acids Research*, **26**, 1848-1850.
6. Lohmann, J.U., Hong, R.L., Hobe, M., Busch, M.A., Parcy, F., Simon, R. and Weigel, D. (2001) A molecular link between stem cell regulation and floral patterning in Arabidopsis. *Cell*, **105**, 793-803.
7. Lamb, R.S., Hill, T.A., Tan, Q.K.G. and Irish, V.F. (2002) Regulation of APETALA3 floral homeotic gene expression by meristem identity genes. *Development*, **129**, 2079-2086.
8. Maizel, A., Busch, M.A., Tanahashi, T., Perkovic, J., Kato, M., Hasebe, M. and Weigel, D. (2005) The floral regulator LEAFY evolves by substitutions in the DNA binding domain. *Science*, **308**, 260-263.
9. Parcy, F., Nilsson, O., Busch, M.A., Lee, I. and Weigel, D. (1998) A genetic framework for floral patterning. *Nature*, **395**, 561-566.
10. Morgan, W.D., Williams, G.T., Morimoto, R.I., Greene, J., Kingston, R.E. and Tjian, R. (1987) 2 Transcriptional Activators, Ccaat-Box-Binding Transcription Factor and Heat-Shock Transcription Factor, Interact with a Human Hsp70-Gene Promoter. *Molecular and Cellular Biology*, **7**, 1129-1138.
11. Hong, R.L., Hamaguchi, L., Busch, M.A. and Weigel, D. (2003) Regulatory elements of the floral homeotic gene AGAMOUS identified by phylogenetic footprinting and shadowing. *Plant Cell*, **15**, 1296-1309.
12. Gomez-Mena, C., de Folter, S., Costa, M.M.R., Angenent, G.C. and Sablowski, R. (2005) Transcriptional program controlled by the floral homeotic gene AGAMOUS during early organogenesis. *Development*, **132**, 429-438.
13. Weinmann, A.S. (2004) Innovation - Novel ChIP-based strategies to uncover transcription factor target genes in the immune system. *Nature Reviews Immunology*, **4**, 381-386.
14. Frohlich, M.W. and Chase, M.W. (2007) After a dozen years of progress the origin of angiosperms is still a great mystery. *Nature*, **450**, 1184-1189.
15. Theissen, G. and Saedler, H. (2001) Plant biology - Floral quartets. *Nature*, **409**, 469-471.
16. Pelaz, S., Ditta, G.S., Baumann, E., Wisman, E. and Yanofsky, M.F. (2000) B and C floral organ identity functions require SEPALLATA MADS-box genes. *Nature*, **405**, 200-203.
17. Honma, T. and Goto, K. (2001) Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. *Nature*, **409**, 525-529.
18. Majka, J. and Speck, C. (2007), *Analytics of Protein-DNA Interactions*, Vol. 104, pp. 13-36.

Figure legends

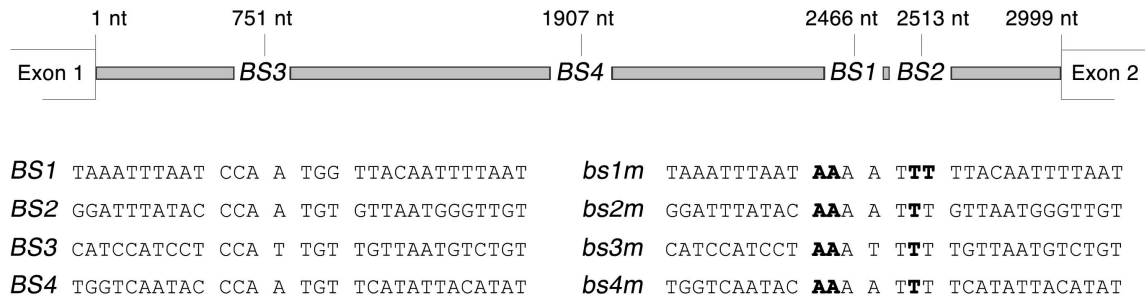


Figure 1. Construction of mutated versions on the AG second intron.

The AG second intron contains four consensus LFY-binding motifs (CCANTG^T_G), BS1-4. Nucleotide positions for the start of each motif are shown relative to the first base of the intron. Mutations (bold type) were introduced to modify three or four bases in each of BS1-4. Mutant sites were combined to generate two mutated versions of the intron, containing mutations in sites BS1 and 2 (*bs12m*) and in sites BS1-4 (*bs1234m*), respectively.

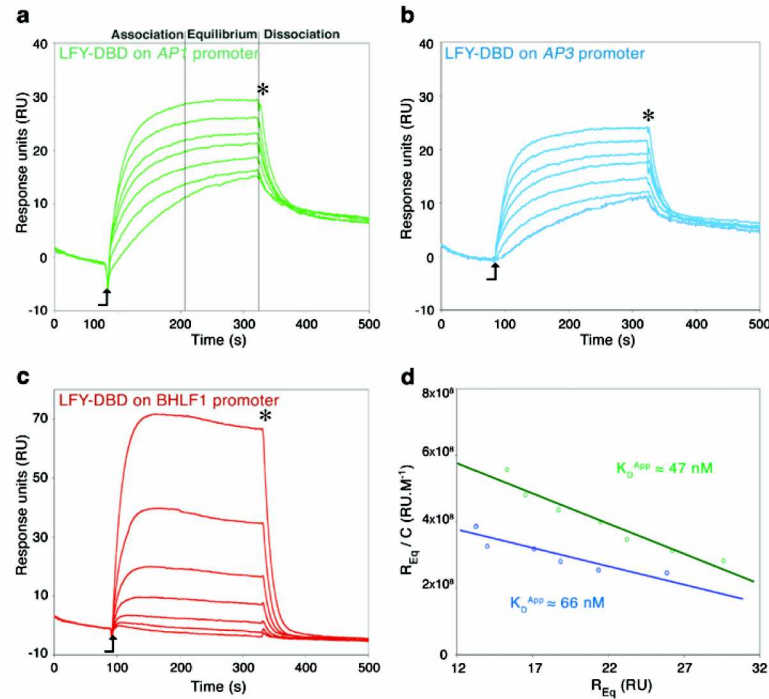


Figure 2. LFY interactions with target and non-target promoters.

(a-c) SPR curves for interactions of the LFY DNA-binding domain (LFY-DBD) with the regulatory regions of the LFY target genes *AP1* (a) and *AP3* (b) and of the non-target gene *BHLF1* from the Epstein-Barr virus (c). The LFY-DBD concentrations used, in descending order of SPR response in arbitrary units (RU), were: 105 nM, 84 nM, 67 nM, 54 nM, 43 nM, 34 nM and 28 nM for target DNA (a and b) and 104 nM, 69 nM, 46 nM, 31 nM, 21 nM, 14 nM and 9.1 nM for non-target DNA (c). The start and end of LFY-DBD injections are marked with an arrow and an asterisk, respectively.

(d) Linear relationships of R_{Eq}/C against R_{Eq} for interactions of LFY-DBD with the promoters of *AP1* (a) and *AP3* (b), where R_{Eq} corresponds to the SPR response (RU) at equilibrium for a given protein concentration (C). The linearity of these plots validates the use of a 1:1 binding model. The K_D^{App} values shown in the figure, derived from the inverse negative of the linear gradients (18), are in good agreement with those shown in Table 2, calculated using SPR analysis software.

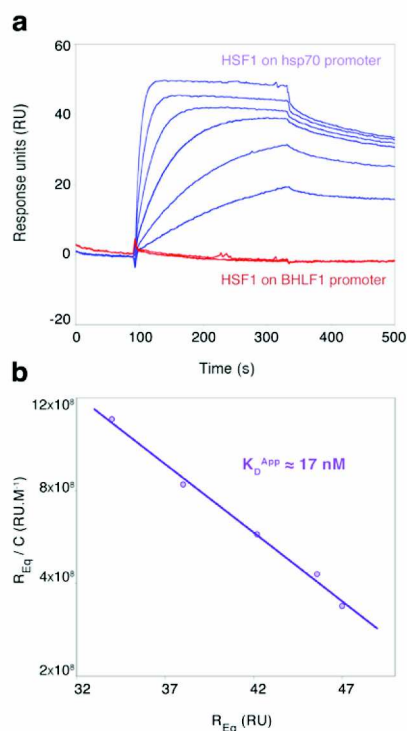


Figure 3. HSF1 interactions with target and non-target promoters.

(a) SPR curves for interaction of HSF1 with the promoter of its target gene, hsp70 (blue traces), and absence of interactions with the non-target, BHLF1 promoter (red traces). HSF1 concentrations used, in descending order of SPR response, were: 155 nM, 104 nM, 69 nM, 46 nM, 31 nM, and 21 nM.

(b) The linear relationship of R_{Eq}/C against R_{Eq} for the interaction between HSF1 and hsp70 validates the use of a 1:1 binding model. A K_D^{App} value shown in the figure, derived from the inverse negative of the linear gradient (18), is in good agreement with that shown in Table 2, calculated using SPR analysis software.

Table 1. PCR amplification of target gene promoters.

The lengths of the amplified fragments are shown before and after cloning in a plasmid vector, which added polylinker-derived sequences to the ends of the biotinylated molecules used in PCR analysis.

Gene	Oligonucleotide sequences	Length of amplified DNA	Length of biotinylated DNA
AP1	5'-CGACTATATGTTCTAAATCCACTTCG 5'-CATTTTTGATCCTTTTTTAAGAACT	2205 bp	2420 bp
AP3	5'-TGAGACCAGATCAAGAGTGCGTGT 5'-TTATTCAAGAAGATGGAAGGTAAT	2873 bp	3088 bp
AG	5'-GTTTCTTCTTCTTCTCGTGCTCTG 5'-GTTAATTAGACCTATTAACCT	2969 bp	3184 bp
hsp70	5'-GTCCCTCCAGTGAATCCCAGAAG 5'-CTAATCTACCTCCTCAATGGTG	2355 bp	2570 bp

Table 2. Kinetic constants for transcription factor interactions with target and non-target genes.

The constants shown were calculated from real time SPR interaction curves using Bio-Eval T100 software (Biacore, Uppsala, Sweden) on the basis of a 1:1 (Langmuir) binding model. k_{off}^{App} and k_{on}^{App} represent the apparent rate constants for dissociation and association, respectively, of DNA-protein complexes. K_D^{App} values were calculated as the ratio $k_{off}^{App}/k_{on}^{App}$ and standard error values for all constants are given. The K_D^{App} values shown are in good agreement with those estimated from the slope of the plots shown in Figs 1d and 2b. χ^2 values of <2 (or of <10 for low affinity interactions with non-target DNA) indicate a good fit of data to the 1:1 binding model, used for the calculation of kinetic constants. Kinetic constants are given for the second intron of *AG* containing no mutations (WT), mutations in two (*bs12m*), and in four (*bs1234m*) LFY-binding sites, respectively, as illustrated in Figure 1.

Transcription factor	DNA regulatory region	k_{off}^{App} (s ⁻¹)	k_{on}^{App} (M ⁻¹ .s ⁻¹)	K_D^{App} (nM)	χ^2
LFY-DBD	<i>AP1</i>	$2.52 \times 10^{-3} \pm 0.5 \times 10^{-4}$	$1.59 \times 10^5 \pm 0.5 \times 10^4$	15.8 ± 0.9	0.249
	<i>AP3</i>	$2.38 \times 10^{-3} \pm 0.2 \times 10^{-5}$	$1.11 \times 10^5 \pm 0.7 \times 10^4$	21.4 ± 2.6	0.200
	<i>hsp70</i>	$5.71 \times 10^{-3} \pm 0.1 \times 10^{-3}$	$4.22 \times 10^3 \pm 0.3 \times 10^3$	~1350	6.73
	<i>BHLF1</i>	$2.96 \times 10^{-1} \pm 0.1 \times 10^{-5}$	$6.76 \times 10^3 \pm 2 \times 10^3$	> 40000	6.2
	<i>AG WT</i>	$7.04 \times 10^{-5} \pm 0.2 \times 10^{-5}$	$1.18 \times 10^5 \pm 2 \times 10^3$	0.6 ± 0.02	1.20
	<i>AG bs12m</i>	$2.38 \times 10^{-3} \pm 0.4 \times 10^{-5}$	$5.68 \times 10^4 \pm 0.9 \times 10^3$	41.9 ± 0.7	0.735
	<i>AG bs1234m</i>	$1.49 \times 10^{-2} \pm 0.5 \times 10^{-3}$	$5.57 \times 10^4 \pm 0.5 \times 10^4$	267.5 ± 30.5	0.654
HSF1	<i>hsp70</i>	$6.48 \times 10^{-4} \pm 0.4 \times 10^{-5}$	$3.03 \times 10^5 \pm 0.3 \times 10^4$	3.5 ± 0.033	0.255